THE UNIVERSITY *of* EDINBURGH

A Single-Cell Study of the Biophysics of Bacteriophage Infection

Jack A. Hocking

Doctor of Philosophy
The University of Edinburgh
February 2024

Abstract

Bacteriophages are the most populous organism on the planet and have huge potential as tools of research in micro- and synthetic biology. They have also more recently gained attention as powerful antimicrobial agents. Our understanding of the underlying processes by which phage infection occurs is in many areas still incomplete. By inspecting the biophysical effect of phage infection on bacteria it may be possible to shed light on some of the mechanisms that still elude us, and point towards areas of further research beneficial to a wide variety of phage scientist. I demonstrate in this thesis that Escherichia coli cells infected by T7 phage exhibit a drop of 5-10% cell volume at the point of infection. If developed further, this volume loss has the potential to be a truly label-free marker of phage infection of a single bacterial cell. This result accords with a mathematical model that predicts loss of volume and solutes by opening of a pore around the phage infection site. The process appears to be stochastic and dependent on the temperature history of the cells, with 36% of unperturbed cells showing measurable losses in volume compared to 58% of cells that have spent a short time at room temperature. Additionally, I present a mathematical model of a single E. coli cell during T4 phage infection wherein the transcription and replication of phage DNA is explicitly dependent on the expression of proteins controlling the processes temporally.

Lay Summary

Bacteriophages (a mix of 'bacteria' and φαγεῖν, *phagein*, - meaning 'to devour') are the most populous organisms on Earth, numbering at approximately 10^{31} organisms in just the oceans alone. While they have been known about and used in molecular biology experiments for about a century, they have recently come into the public eye as potent therapeutic agents. The rise of anti-microbial resistance among infectious diseases has been listed as an extreme health concern by the World Health Organisation, and bacteriophages (or 'phages' for short) represent an opportunity to overcome it. Despite a century of study as viruses of microbes and this recent push into therapeutic applications the fundamental physical process underpinning phage infection are not well understood. It is known, for example, that the phage must ensure its DNA is completely and safely inside a potential bacterial host, however the exact mechanism by which this occurs differs by phage strain and has no unified physical understanding.

In this work I present a measurement of host (in this case *E. coli*) volume over time as the cells are infected by phage T7. The bacteria appear to lose between 5% and 10% of their volume on infection much in the same way that a bike tire might lose some of its volume when introduced to a nail. However, unlike the tire the cells will then recover this lost volume before dying. The mechanism of recovery appears to be osmotic in nature, wherein the cell loses some of the internal molecules that encourage an inflating flow of water into the cell and gradually accumulates more. Imagine an inflatable skydancer. If one were to bring a sharp implement to him he may sag slightly before the ensuing hole is taped over. Thus the cell too loses volume before the hole is repaired and the normal pumping systems can bring it up to it's original volume.

The measurements of *E. coli* volume span two data sets. One gathered in cells that had a roughly half-hour stint at room temperature and one in cells that were held at 37°C (optimal growth temperature) for the entire duration of the

experiment. In the room temperature cells it was seen that they would exhibit the loss of volume (termed 'dip') about 1.6 times more than the happy 37°C cells, and that those dips had approximately twice the duration. The current hypothesis for this is that the *E. coli* membrane starts to become a gel at room temperature, and so is less able to move around. In this way the membrane's ability to passively heal around the infection site is impeded, and the volume dip should be longer. This work is supported by a mathematical model of cell pressure and solute content that predicts the variation in dip length as a function of the time the hole is open for.

Acknowledgements

I would like to thank the Soft Matter and Functional Interfaces CDT for accepting me and allowing me to begin my journey and introduced me to a vibrant cohort of peers, many of which have become good friends. My supervisors Aidan and Teuta I have found exceptional in their support academically and personally, and I would simply not have been able to complete my studies without them. I could not have asked for a better pair of mentors. I was fortunate enough to live with my parents for the duration of my studies, without them things (especially lockdown) would have been significantly harder and for them I have the utmost love and care.

I would also like to thank my partner Cat, in whom I find a wonderful study partner and constant source of joy; and my dear friends Lucas and Veronica for many thoughtful conversations and discussions, and for putting up with my various complaints and ills. And I extend my gratitude to my Kung Fu Sifu, Conrad Barrett-Freeman, and his school, who have kept me sane over this long, strange, journey.

And lastly, I would like to thank the Delta Star discord server who have, despite their best efforts, collectively and significantly improved my time during my project.

Contents

List of Figures

List of Tables

Chapter 1

Background Theory

1.1 Historic Phage Study

The history of bacteriophages as both therapeutic agents and research tools is quite unique and I would like to spend some time on it before introducing the topic of my thesis proper. The first report of phage activity is often cited to be by Frederick Twort in 1915[132]. However, he attributed their action to an enzyme excreted by bacteria and named them the 'bacteriolytic agent'. In 1917 they were independently discovered by Felix d'Herelle who correctly identified them as 'invisible microbes antagonistic toward dysentric bacilli' and proceeded to immunize rabbits against *Shiga dysenteriae* and cure dysentery in a human[106]. D'Herelle recognised the potential for his *bactériophage* in medical therapy and travelled the world setting up clinics for treatment of bacterial infections. However, the work of D'Herelle and collaborators often failed to meet accepted standards of scientific rigour at the time, leading to a fast decline in support from the wider community. This, alongside the advent of penicillin, which was far easier to use than phages, led to phage therapy falling out of favour in the West. It would however continue in the East, which posed its own problem in that it bestowed upon phage therapy a so-called 'Soviet Taint'[125]. The Cold War pushed phage therapy away from western medicine as it was seen as a 'communist science' and thus subject to the negative propaganda that accompanied the label (not helped by the frankly pseudo-scientific Lysenkoism that was popular at the time)[35][122]. As such, our understanding of phages and their capabilities in medicine is rather limited compared to that of ex-Soviet countries - particularly

Georgia which hosts the Eliava Phage Therapy Institute in Tbilisi[102].

However while their potential as therapeutic agents is only now being fully realised in the West, phages have long enjoyed a place by the side of the micro- and molecular biologist. Notably the American Phage Group (APG), formed by Max Delbrück and Salvador Luria, were responsible for a number of foundational experiments in the 1940s and 50s with phage. Shortly after forming the group, the founding pair performed their well-known Fluctuation Experiment[79], later dubbed the Luria-Delbrück Experiment. The aim of this was to provide an answer to the argument prevalent at the time between Lamarckian and Darwinian evolution theory. The former states that beneficial mutations should arise in direct response to environmental hostility[15], while the latter predicts that these mutations arise randomly and subsequently the environmental conditions 'select' cells that have those mutations[31] – i.e. bacteria not fortunate enough to have mutated in a beneficial way will die off. The Darwinian theory thus implies a greater variation in the size of resistant bacteria populations after phages (the environmental hostility) have been added since resistance is inherited leading to either small populations (the resistance is recent) or large populations (the resistance is old) as can be seen in figure 1.1.

This experiment marks the first major result for phages as a tool for biological research[113]. Further experiments from the Phage Group - notably the Hershey-Chase experiment - took phage research further and helped spur the creation of the field of molecular biology (indeed, Delbrück is often cited as one of the field's founders). Prior to Hershey-Chase, it was assumed that the genetic information of cells was carried by proteins and DNA was simply a repeating sequence of nucleotides that served some other function [34]. In their work, Hershey and Chase used radioactive phosphorous-32 to label the DNA of the T2 bacteriophage - since phosphorous is found in DNA - and sulfur-35 to label the protein shell. They showed that during infection the sulfur-labelled phage proteins remained on the outside while the phosphorous-labelled DNA was injected into the bacteria. Therefore, the previous assumption of protein-carried information was incorrect, and it must be DNA that instead contains genetic information. They also concluded there must be some mechanism by which the DNA is injected into the cell, the exact mechanics of which remains poorly understood even today [46].

Recently (within the writing period) there has been a shift in the taxonomic description of many phage species. Of particular interest is the abolition of the order *Caudovirales* and its families *myoviridae*, *podoviridae*, and *siphoviridae*. As

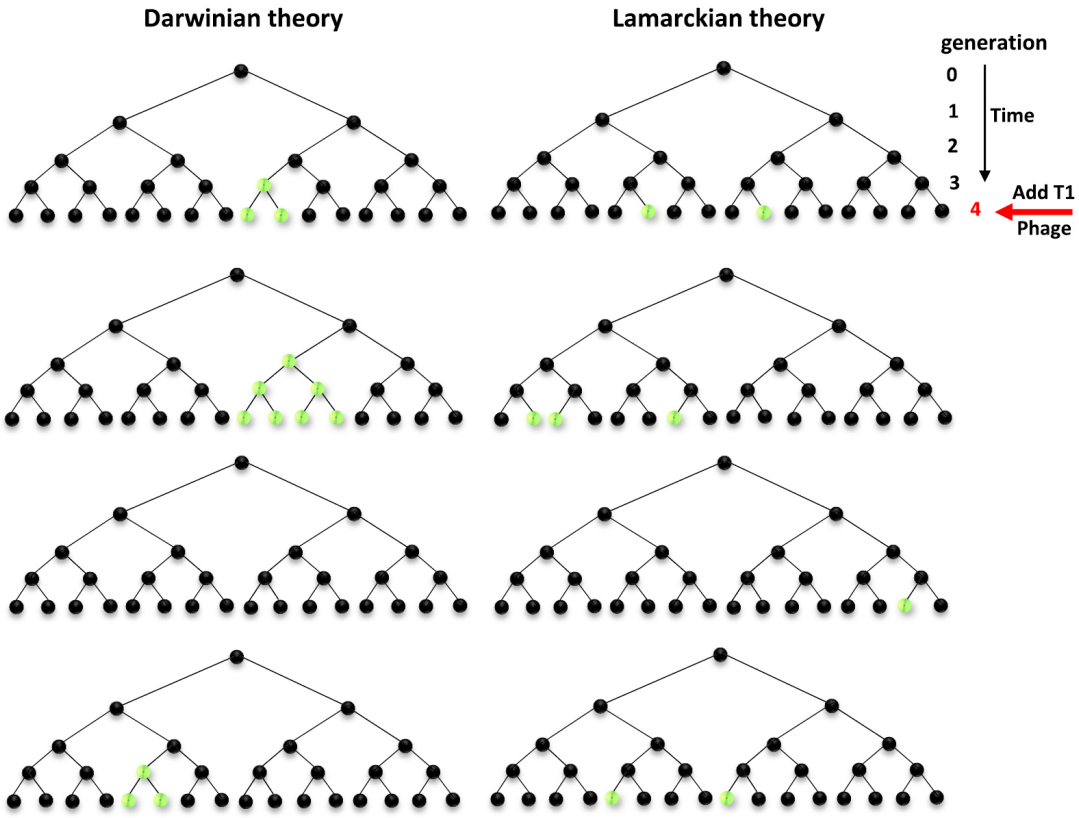


Figure 1.1 *A cartoon showing the expected results of the Luria-Delbrück experiment given the two evolutionary models. Resistant mutants are shown in green. In the Darwinian model it assumed that resistance does not only form when the bacteria are exposed to phage, rather any bacterial population has a chance at becoming resistant to future phage additions at any stage through random mutation. Luria and Delbrück found that their Fluctuation Experiment followed the Darwinian model, and thus that the idea of natural selection provides a better fit for bacterial evolution in hostile environments, rather than the direct response postulated by Lamarck. One can see in this image the variation in numbers of resistant populations is much greater in the Darwinian model as resistance is inherited from previous generations rather than acquired only at the moment of exposure to phages and it is more likely for non or all of the populations to be resistant than in the Lamarckian model. Image taken from[49]*

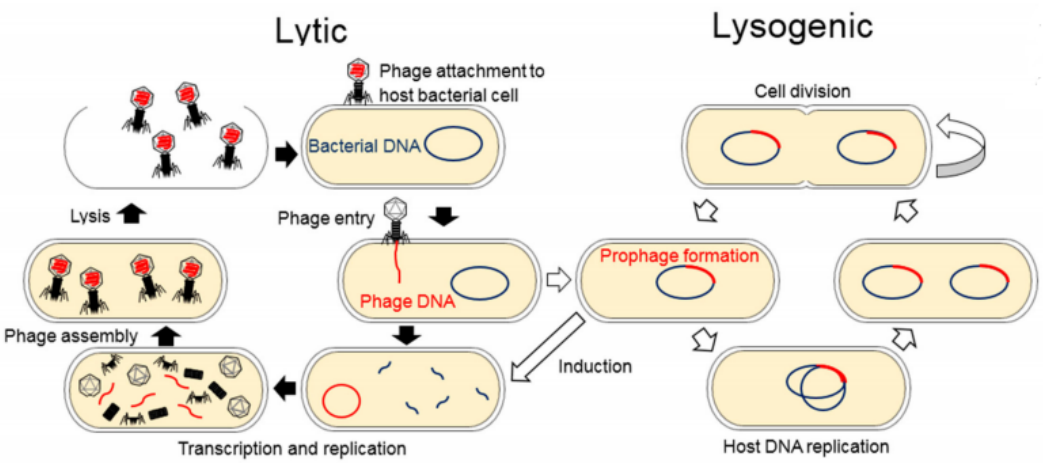


Figure 1.2 *Cartoon depicting the lytic pathway with the optional lysogenic pathway available to temperate phage. Phage that do not engage in lysogeny are known as lytic or obligately lytic for emphasis. Image taken from[126]*

seen in table 1.1 these are the families to which the T-series phages popularised by the APG belong. This shift comes from a retrospective look at phage classification, wherein the phage families had become polyphyletic. The new families, *straboviridae*, *autographiviridae*, and *demerecviridae* describe phages more strictly than simple appearance. For example, *podoviridae* used to define phages that were 'T7-like' in appearance and function (but not necessarily strongly linked), but *autographiviridae* defines phages that encode their own RNA polymerases. The terms 'myophage', 'podophage', and 'siphophage' are still in use as polyphyletic descriptors of phage shape.

1.2 Bacteriophage Infections In General

While each taxonomic family is quite different in terms of morphology, host specificity, and behaviour, the infection process can in general be boiled down to four steps: adsorption, DNA-injection, host takeover, and finally lysis of the cell. There are two distinct pathways by which phages may replicate: lytic and lysogenic (figure 1.2).

With these two pathways comes two groups of phages that are generally distinct. Obligately lytic phages will act as above, simply infecting and lysing a population of cells until either the population has collapsed or some steady-state of division and death has been reached as both host and virus competitively mutate around

each other[21]. Temperate phages are able to use the lysogenic pathway in which the phage DNA is incorporated into and replicated with the host genome without the creation of phage progeny and without immediate lysis of the host. After some external condition is met the temperate phages will leave the lysogenic state and return to the lytic pathway, lysing the cell and releasing progeny into the environment. Some temperate phages will integrate their DNA with the host through random transposition (the phage DNA acts as a transposon) or site-specific recombination (the DNA integrates at a specific gene site) in the cases of *E. coli* phages Mu and λ respectively; whereas other phage DNA will remain as separate either linear or circular DNA molecules (plasmids) as in phages N15 and P1[18]. Temperate phages can be more likened to farmers than pesticides, as they are able to 'care' for their host bacteria by reducing populations of competing strains and providing host-beneficial genes via their internalised genomes[12]. Temperate phages can also increase a population's resistance to antibiotics and potentially other phage[103]. Additionally, some temperate phages have been shown to increase the virulence of some pathogens and make treating infections in plants or humans much more difficult[2][115][82]. All this combined means that temperate phages are not ideal for therapeutic research. I will focus more on lytic phages as these are the phages that I have used.

1.2.1 Adsorption of Phage to the Cellular Membrane

After a phage has diffused through a host population long enough to successfully collide with a host it must adsorb to the membrane. When this was first studied there appeared to be a paradox in this process in that far more successful adsorption events happened than apparent collisions predicted from Brownian statistics[123]. The first solution to this paradox was presented in 1949 when it was noted that there appeared to be some "steric fitting of the elements and of the formation of a weak bond between virus and host"[4]. That is, the number of collisions alone could not account for the number of successful infections if there was not some mechanism by which the phage could 'stick' to bacteria. This would be confirmed later on in a comprehensive study of the temperate phage λ wherein this model of a steric bond provided an adequate explanation for the apparent paradoxical behaviour.

Of course we have since significantly increased our understanding of the underlying mechanisms and now understand that this 'steric fitting' is the result of

Phage	Family	Family (Old)	Receptor
T1	*Demerecviridae*	*Siphoviridae*	TonA (FhuA) and TonB
T2	*Straboviridae*	*Myoviridae*	OmpF and Fadl
T3	*Autographiviridae*	*Podoviridae*	Glucosyl-α-1,3-glucose terminus of LPS
T4	*Straboviridae*	*Myoviridae*	OmpC and LPS
T5	*Demerecviridae*	*Siphoviridae*	Polymannose in O-antigen and FhuA
T6	*Straboviridae*	*Myoviridae*	Outer membrane protein Tsx
T7	*Autographiviridae*	*Podoviridae*	LPS

Table 1.1 *Table reproduced from[9] showing the receptors for each of the T-series phages found in their host E. coli. The table illustrates the wide variety in receptor across and within phage family, while also showing that different phage can utilise the same receptor, e.g. T1 and T5 both use the TonA (FhuA) protein. Phage families have been changed in accordance with recent changes to taxonomy as ratified by the International Committee on Taxonomy of Viruses wherein the order caudovirales and families myoviridae, podoviridae, and siphoviridae were abolished[133]*

the source of phage host-specificity and the key to phage adsorption - binding receptors. Each phage has evolved to recognise certain membrane features, often porins or other populous membrane-bound protein, in the outer membrane as receptors to bind to. An example of the range in phage receptors in the host bacteria *E. coli* can be seen in table 1.1.

Proteins responsible for binding to phage receptors are found on the tails of the phages and the binding event generally induces a conformational change in the phages which leads to phage DNA injection into the host cell. The types of binding and conformational changes are as different between phages strains as the receptors themselves. For example, in T4 the phage will use protein structures known as 'long tail-fibres' (LTF) to reversibly stick to lipopolysaccharides (LPS) on the *E. coli* membrane. The LTF bind and unbind from LPS on the membrane in a random process allowing the T4 virion to travel across the membrane in a diffusive process often referred to as walking across the membrane until it irreversibly adsorbs via its short tail fibres [51]. In general phage-host specificity is determined by binding viability between phage tail fibres and host membrane features; the specific requirements of T4 are discussed in Section 1.3.2. T4 then undergoes a conformational change as its tail sheath contracts and rotates as it's thin inner tube spans the periplasmic space in three steps. First, it mechanically pierces the outer membrane; second, it digests the peptidoglycan layer with a lysozyme tip; and third, the cytoplasmic membrane bulges up locally to meet and fuse with the tip[81]. In another example, T7 - the phage on which I focus

in my thesis - will similarly diffuse across the membrane (detailed in Section 1.3.1) but on irreversible binding initiates a complex injection process involving a protein structure implanted and unfolded in the host[116].

The irreversible nature of the final binding raises the question as to whether there are other requirements. For optimal population growth the phage should minimize both self-competition and non-productive infections. A high phage to host ratio would naturally lead to many phage infecting a single host, leading to a reduction of free phage without additional infections and potentially an overwhelming of the host - killing it outright. In T4, for example, it was originally believed that the proton motive force (PMF) - the potential of protons across the membrane driving ATP synthesis and flagella motor rotation - is necessary for the phage to bind to their host. This assumption was supported by the demonstration by Grinius *et al* that infection of de-energised *E. coli* cells by T4 leads to inactive phage particles, and additionally that a T4 infection causes the membrane potential - and subsequently PMF - to collapse[40]. However, it has since been shown that it is only required for membrane puncture and DNA injection - so the aforementioned inactive phage particles are inactive due to being irreversibly stuck to something they cannot infect[71]. This lack of infection then is potentially indicative of a mechanism by which the phage may avoid overwhelming their hosts, rather than any sort of avoidance of losing phage to non-productive adsorption.

Indeed, there is a phenomenon known to occur in T-even phage, and other phage with large genomes, known as lysis-from-without (LWO). LWO occurs when there are very high multiplicities of infection (MOI - the phage to bacteria ratio) and a cell lyses from phage pressure rather than complete infection[1]. If one considers a cell as a rather more complex car tyre, one can imaging the effect of a single phage as that of driving over a single upright nail. This is not likely to severely compromise the tyre, however an entire box of nails would have rather dire consequences. Similarly, a cell infected by an overwhelming number of suitable phages will simply be ripped apart[24]. This process has been proposed to be similar to altruistic cell death, as in order for LWO to occur the number of free phages must be much higher than that of the available hosts - since wiping out the host also means wiping out the virus it is beneficial to reduce the number of phages in the area[29]. Interestingly, LWO is not a common phenomenon and phages with smaller genomes, including T7, will simply overwhelm the host's ability for macro-molecular synthesis leading to simple host death[91]. Very high

MOIs indicate a large population imbalance so LWO may be beneficial to these phage to prevent non-productive adsorptions as binding to cell debris ought to be less efficient. Infection by LWO-able phage confers a phage-product resistance to LWO, via phage gene products sp and *imm*[1].

1.2.2 DNA Ejection

Much work has been done since the revolutionary Hershey-Chase experiment over 60 years ago. However, the exact mechanism by which phage DNA is injected into the host cytoplasm is still unclear. There is no one mechanism that governs the translocation of phage DNA for all of phage-kind, different phages must inherently rely on different processes which themselves are complex.

For example, T4 has one of the fastest DNA ejections at around 4000 base pairs (bp) per second, which has been suggested to be driven by mechanical forces arising from the near-crystalline packaging of the phage DNA within the capsid. The injection of T4 DNA happens as one process, unlike another example the *B. subtilis* phage ϕ29 which injects around 65% of its DNA - likely using internal packing pressure as above - before a molecular machine is built from that portion of DNA to bring the rest of the genome inside[38]. T7 is again different, it uses a capsid-contained protein complex that unfolds to create a channel into the cytoplasm through which a small section of genome is inserted before the rest of the genome can be grabbed by the host RNA polymerase (RNAP) molecules[90].

In vivo DNA ejection is somewhat predictably more complicated as there are now two compartments to consider as the bacterial cytoplasm must be taken into account. This presents an additional pressure as the cell maintains its inflation through osmotic forces (discussed in Section 1.4), as well as introducing additional biological parameters. In infections with phage T4, bacterial populations with higher growth rates prior to infection lead to high phage growth rates post infection, as shown by Nabergoj *et al*[96]. Using chemostat dilution rate as a proxy to growth rate they saw that as growth rate increases the burst size is increased and the latent period decreased, meaning more phages are produced per lysis event and the time between infection and lysis is reduced. Interestingly the adsorption rate of T4 to *E. coli* was shown to decrease with increasing growth rate. They theorised that this was because larger cells should have a lesser concentration of OmpC in the membrane.

I have identified, broadly speaking, three competing hypotheses regarding the exact mechanism of DNA internalisation from phage to bacterium. Two of these models assume that injection is purely driven by pressure differences, and the third takes into account the contents of the cytoplasm into which the DNA is being injected. As the pressure models rely on a reasonable understanding of osmotic pressure, I have described these models in Section 1.4.2. after a treatment of the background theory of osmotic pressure in cells.

1.2.3 Host Takeover

Once phage contents have been trans-located to the host cytoplasm the process of hijacking the bacterial biomachinery in order to produce new phage progeny begins. Understanding the essentials of phage DNA transcription and translation is not necessarily a requirement for the main result presented in this thesis. However, they are part of the infection process and as such may be impacted by the loss of cell pressure I present. Additionally, I present a computational model in Chapter 6 that aims to track the host takeover of *E. coli* by T4 with regards to the intracellular processes that take place. In Chapter 6 I include a small section of background that pertains more to the model than to the experimental results presented in Chapter 5.

Broadly speaking phage gene expression is a temporally controlled process with distinct stages often labelled as 'Early', 'Middle', and 'Late', each with their own class of promoter - though some phages can lack the 'Middle' phase. The naming convention here refers to the temporal development of the phage protein production. Proteins produced from early promoters during the early phase work to bring the system into the middle phase where middle promoters likewise work towards the late phase. Genome transcription in phage infections is primarily classed by the presence of endogenous RNAP holoenzymes. Phages such as T4 and λ do not encode their own RNAP and so must adapt the host RNAP to transcribe their DNA. Phages such as T7 and T3 do encode their own RNAP, and so generally rely on host RNAP to express the phage-encoded RNAP that lies under an Early promoter. The host RNAP is then inhibited by a phage protein and the phage RNAP continues the process until lysis[144].

	T4	T7
Family	*Straboviridae*	*Autographiviridae*
Receptor	OmpC & LPS	LPS only
RNAP Source	Host	Phage
Latent Period*	20-25 min	13-17 min
Burst Size*	150	260
Genome Length	170 kbp	40 kbp

Table 1.2 *Quick reference for phages T4 and T7. *Numbers reported for phages infecting an exponentially growing bulk population at 37°C[23].*

1.3 Bacteriophage Strains Used in this Thesis

The phage strains used in this thesis are T1[1], T4[2], and T7[3]. The main result of this thesis in Chapter 5 was observed with T7. It was believed that we had stocks of T4 (not T1 masquerading as T4) and so much of my work prior to obtaining that result concerned T4, including Chapter 6. The mistake with T1 was found with cryo-EM imaging performed by Ryan Morris, and the labs were quickly sterilized. Some cell sizing experiments were performed with T1, however an error in the experimental process rendered the data unusable and T1 is not discussed in this thesis beyond Section 1.3.3. A quick-reference for both T4 and T7 can be seen in table 1.2.

1.3.1 The T7 Bacteriophage

T7 is the prototype of the 'T7 group', a classification of T7-like phages thought to be obligately lytic[89]. It has a roughly 60 nm diameter capsid and a genome length of 40 kbp. It is one of the 'T-series' phages (coliphages popularised by Delbrück *et al*) and was a member of the *podoviridae*, a polyphyletic family of phages characterised by short, non-contractile tails. T7 is now part of the monophyletic family *autographiviridae*, referring to the 'auto-graphein' or 'self-transcribing' nature of those phages which encode their own RNAP[11].

T7 adsorption is dependent on interaction between LPS and the tail fibres. While it has been reported that FhuA and TonB serve as receptors for T7[108] it is generally held that T7 relies on rLPS on the bacterial membrane as is noted

[1]Accidentally, see Section 1.3.3.
[2]Briefly, after finding the accident.
[3]For the majority of my experiments.

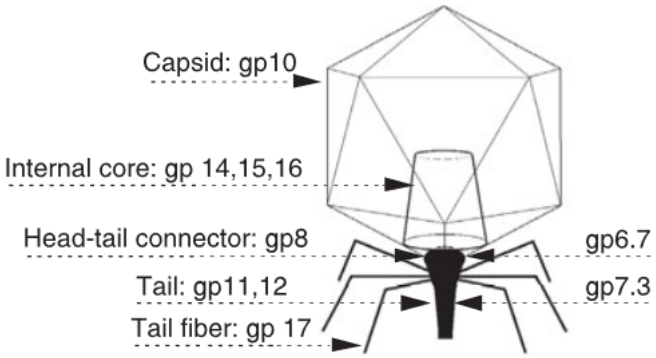


Figure 1.3 *A schematic of T7 with protein structure labelled. The packed genome takes up about 40% of the capsid volume. During infection the internal core formed of gp14, gp15, and gp16 is used to form a channel into the cell through which the DNA is internalised. Image taken from[89]*

in table 1.1. T7 diffuses along a bacterial membrane in a process which has been described as 'walking' or 'hopping' wherein some of its tail fibres will bind reversibly to the membrane and then unbind, dissociating phage from membrane. Injection of DNA is dependent on the phage being well aligned with the membrane, which occurs when all the tail fibres bind, which brings this hopping diffusion to an end as in figure 1.4.

In plasmolysed cells it was found by Bayer in 1968 that there were areas of adhesion between the *E. coli* membrane and cell wall, now commonly referred to as 'Bayer patches'. In the same work bacteriophages T1, T5, and T7 could be seen via electron microscopy to be adsorbed *only* to these points on the cell, indicating that they represent primary or very efficient binding sites for the phages[8]. It has been suggested that the phages in this case could actually be responsible for these patches, and that a trans-periplasmic channel formed during infection could facilitate their formation[90]. The efficiency and simplicity of binding to LPS has caused some small issues for the study of T7 and T7-like phages as the tail fibres can quite easily adapt to other strains of bacteria as the required mutations are quite minor. While this is not an issue if one is certain of one's strain, instances like the T7-like *Yersinia pestis* phage $\phi A1122$ being described in published work without reference to its extended host-range nature or instances of mistaken identity (particularly in T7) are quite common, especially in the past prior to quick genome sequencing[141].

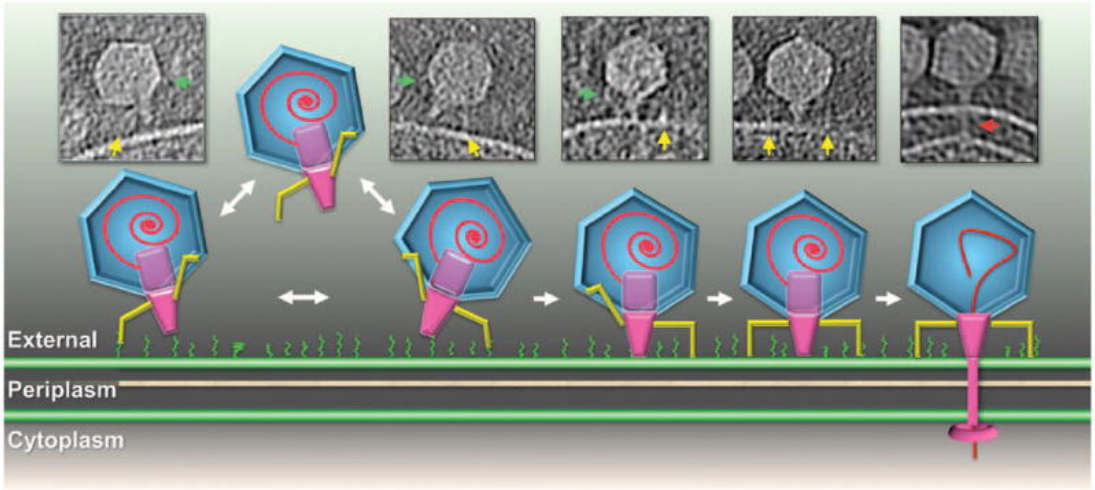


Figure 1.4 *A schematic of the T7 reversible and irreversible binding to the E. coli membrane (below) accompanied by tomograms of the process (above). In the above images yellow arrows denote fibres bound to the membrane and green arrows fibres that are free. Those that bind are rotated downwards, and the final irreversible step occurs when all fibres are bound together. Image taken from[50].*

The T7 injection mechanism complex and involves significant structural changes as it releases its payload into the host. The internal core proteins highlighted in figures 1.3 and 1.4 are ejected from the capsid forming a cross-periplasm channel through which DNA is internalised. It has been found that gp14 remains in the outer membrane of the infected cell while gp15 and gp16 form the channel into the membrane. Gp16 appears to be a versatile protein as the N-terminal domain bound lytic transglycosylase is responsible for breaking down the cell wall[86]. In addition, gp16 appears to be responsible for termination of the initial step of DNA injection wherein about 1 kbp of the genome is internalised[17].

Subsequent complete genome entry takes around 10 minutes as it is internalised by first *E. coli* RNAP molecules and then T7 RNAP. The long inernalisation time allows an early T7 gene product - gp0.3, later named Overcome Classical Restriction (Ocr)[68] - to be produced in sufficient numbers to inactivate host restriction enzymes that would typically cleave the phage DNA as part of the cell immune system. Moffat and Studier found that production of this sufficient number of Ocr took about 6 minutes, but that prior to this the DNA was not susceptible to restriction by enzymes that were not being restricted however the mechanism by which this DNA escapes restriction is unknown[88]. The mechanism of Ocr is one of mimicry, wherein it appears to host enzymes to be DNA allowing it to effectively replace T7 DNA in the restriction system[134].

A recent study by Ye *et al* showed that Ocr also binds to host RNAP, revealing a versatility in this protein allowing it to interfere with host biosynthesis beyond simply protecting the phage genome[146]. This interference is avoided by T7 as it encodes its own RNAP (unlike T4 which uses host enzymes). The T7 genome is small and encodes only 57 proteins (as of latest Uniprot proteomic entry), including a proprietary RNAP which means host RNAP can simply be inhibited rather and thus less proteins are required for the infection compared to phages that adapt the RNAP (such as T4)[85] .

T7 gene products are transcribed in a temporal process comprising an early, middle, and late phase with the respective proteins often referred to as classes I (gp0.3-1.3), II (gp1.6-6.3), and III (gp6.5-19.5). Early gene products are transcribed via host RNAP enzymes and include T7 RNAP which transcribes middle and late proteins. The early and middle periods are responsible for production of proteins that encourage phage RNA production and inhibit host biosynthesis, while the late period is reserved for production of proteins required for both new phage progeny and host lysis. Inhibition of *E. coli* RNAP continues past Ocr production with the early product gp2, which efficiently inhibits the enzymes halting early transcription and pushing the system into the middle phase[114]. Additionally the class II gene product gp5.7 has been reported to be (while overall nonessential) required for optimal phage growth in *E. coli* by providing an additional inhibitor for host RNAP[127].

Several genes in the early and middle phases have been investigated for therapeutic purposes - often as part of studies looking for novel antimicrobial agents - however many remain un-characterised. T7 encodes for several genes classed as growth inhibitors for *E. coli*, each acting in different ways and conferring different benefits to the phage. Gp0.4, a very early class I product, has been found to inhibit the host 'filamenting temperature-sensitive mutant Z' (FtsZ) protein which is essential for cell division[61]. Expression of this protein is lethal for the host, and would result in elongated cells were it not for other phage proteins that more completely halt cell growth. This offers a benefit to the phages as if multiple phage genomes are within one cell, inhibiting the division of that cell hinders the overall phage population, therefore driving competition in phages. As gp0.4 is active for about 4 minutes (one fifth of the optimal *E. coli* division cycle) 20% of the infected population of bacteria will have divided in the absence of gp0.4. Half of these daughters would not give rise to any T7 progeny as they would not contain the phage genome. Thus FtsZ inhibition by gp0.4 can increase progeny

of a phage by roughly 10%. Additional inhibition of growth arises from class I products gp0.6 and gp1.6, and class II products gp3.8, gp4.3, and gp5.3, however the underlying mechanisms by which this occurs are still unclear[93].

The T7 lysozyme, gp3.5, is responsible for regulating class II and class III expression. Gp3.5 binds to the T7 RNAP which decreases its affinity for T7 promoters by destabalizing initial transcription complexes (a complex being an RNAP molecule bound to DNA)[87]. Because class II promoters are comparatively weaker than their class III counterparts they are more greatly effected and class III expression increases. Gp3.5 only interferes with the initial complex formation, once the transcription complex has begun elongating along the DNA it is immune to the activity of the lysozyme[52].

Class III gene products are those involved with the finalisation of phage progeny - DNA packaging motors, phage assembly proteins, and lysis proteins. The T7 lysis process is not well understood. It does not appear to be dependent on expression of any one protein, rather a complex process involving different lysozyme and holin-type proteins that can be up or down regulated as necessary. It has been found that a T7 mutant that cannot express lysozyme will demonstrate a delay in lysis time[124]; and similarly one that has a deletion of the holin protein gp17.5 will also delay a lysis time, but will additionally have an increased production of other class III mRNAs, suggesting a compensatory up-regulation of other proteins involved in the process[143].

1.3.2 The T4 Bacteriophage

The T4 bacteriophage is a member of the *straboviridae* (a family defined by the fact that its members are 'T4-like'[133]). T4 consists of a 126 nm long and 86 nm wide prolate icosohedral capsid - containing the phage's 168 kbp genome[85] packed to a density of about 500 mg/mL - attached to a long contractile tail sheath. At the connection between capsid and tail there are several fibres forming a collar known as whiskers which serve a dual purpose. First, they keep the T4 long tail fibres (LTF, as introduced in Section 1.2.1) in place while the phage is in free diffusion; and second, they act as a sensing device - when the phage is near a potential host they will unfold allowing the LTF to adsorb to the LPS or OmpC on the cell membrane[19][54].

T4 has a fairly narrow host range - infecting some species of *Escherichia* and

Shigella - which is determined by its ability to bind to either OmpC or a mutant of LPS that contains one or two glucose residues. T4 can infect cells in an OmpC-dependent process or an OmpC-independent process [137]. In K12 *E. coli* OmpC is known as the primary receptor; if it is present it will be prioritized over the secondary receptor LPS. A key factor as to whether LPS may be used as a receptor is their 'smoothness'. LPS that do not posses a saccharide side-chain known as an O-antigen are known as rough LPS (rLPS) and those that do are known as smooth LPS (sLPS)[74]. The O-antigen itself is part of a cell's immune system and prevents surface LPS from becoming hydrophobic and thus protecting the cell from hydrophobic antimicrobial agents[131]; they are also used widely as phage receptors for several other strains[13]. Through structural docking analysis it has been shown that the distal tip of the LTF is similar in size to the surface cavity formed by OmpC ($\sim$ 25 Å) and the tip of the LTF sits in this cavity, likely interacting with the extracellular loops of the protein pore[6][7]. Despite LPS and OmpC being quite different structurally the amino acids within the distal region of the LTF that bind to them are similar, the patches formed on the bottom and sides of the trimeric distal tip function for both rLPS and OmpC interaction[128]. The loose binding with these sites allows T4 to travel across the membrane similarly to T7, however it does not dissociate from the membrane during the process - possibly as a result of having longer tail fibres. While there are generally many copies of OmpC (and rLPS) on the host, T4 prefers to infect either at a pole or at an impending division site[30][44]. As an aside, this may explain the result of finding a reduced adsorption with larger cells as one would expect the OmpC proteins to be more spread out thus reducing the probability of being able to bind to several at once. Once at least three tail fibres bind to the receptors in the membrane, the short tail fibres that lie folded at the distal end of T4's tail will unfold and bind irreversibly with the outer oligosaccharide core of LPS in the membrane[98][137].

After the short tail fibres have bound, the phage must puncture the membrane and create a channel by which its DNA may pass directly into the cytoplasm (thus bypassing any periplasm-bound immune system, e.g. nucleases[76]). Recent models have shown that it is likely that the T4 virion rotates as the tail contracts, and in combination with a lysozyme tip to digest the peptidoglycan layer[58] this is what allows it to bore into and span the periplasmic space[37]. The inner membrane bulges up to meet the tail tube and a complete channel connecting the inner and outer membrane is formed (as seen in figure 1.6)[51]. At this point the previous question of models of DNA injection becomes relevant, as it

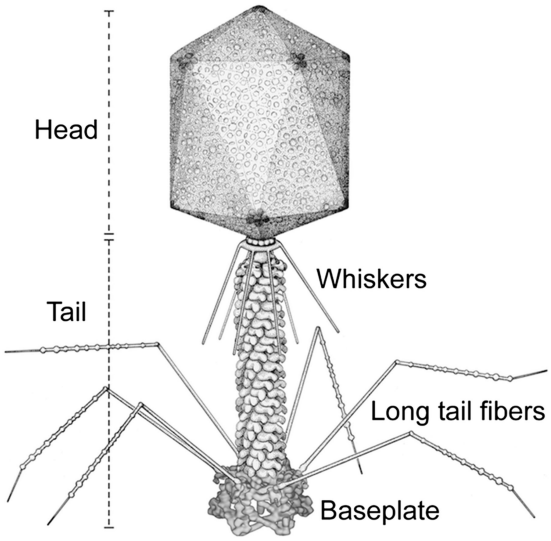


Figure 1.5 *3D cartoon depicting a T4 virion. Visible on the capsid are two of the three proteins comprising the elongated icosohedral structure: gp23, which forms the ring-like structures visible on the faces; and gp24, which forms pentamers at the vertices of the capsid. At the base of the capsid gp20 can be seen, forming the dodecameric portal vertex by which DNA may leave the head[109]. The helical structure of the tail is visible, formed by gp18, and the long tail fibers are in their unfolded state. Image taken from[145].*

is currently unknown exactly how the T4 payload is brought into the host. It is known that in assays of Rb^+ ion efflux (used in analogy to potassium and magnesium flux) T4 induces a transient leakage of intracellular ion content at the point of infection, likely due to being flushed out of, or around, the hole along with water as turgor pressure is lost[69]. I present a detailed overview of the subsequent T4 host takeover in Chapter 6.

1.3.3 The T1 Bacteriophage

If you are using T1 in your lab, pretty soon everyone in the building will be using T1.

Apocrypha

T1 is a member of the *Demerecviridae* family; it is siphoviridic, meaning it has an icosohedral head and a long flexible and non-contractile tail. It has a genome length of approximately 50 kbp, encoding some 77 different genes, only 27 of

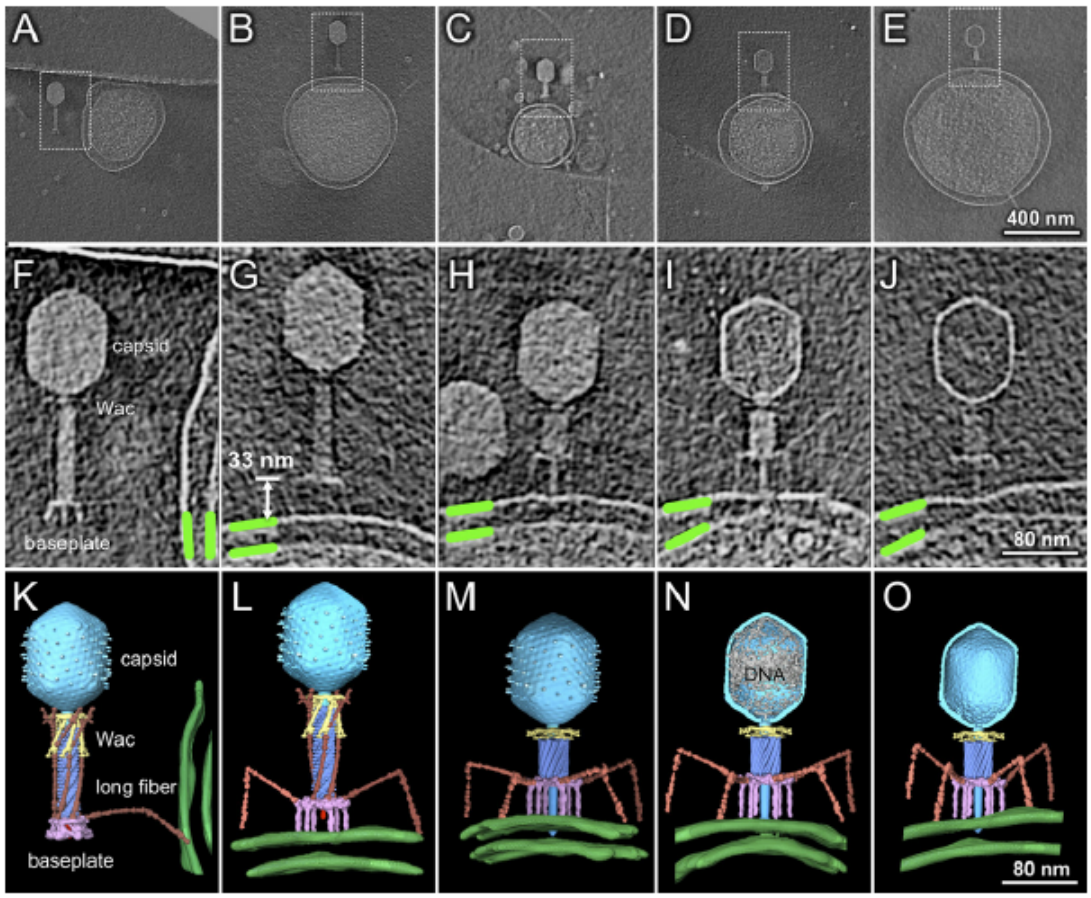
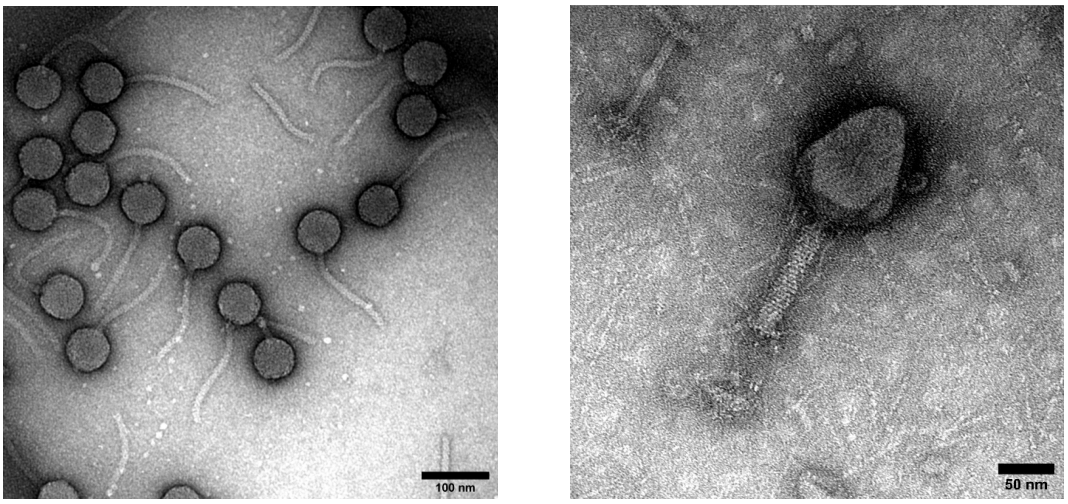


Figure 1.6 ***A-J**, 3D tomograms shown as central slices of the stages of infection of E. coli by a T4 phage. In **F**, a single long tail fiber can be seen attached to the surface of the cell before the phage is brought into contact. In **J** the phage's tail has contracted forcing the tube inside the bacteria. The inner membrane can be seen bulging upwards, possibly the result of an electrostatic attraction due to the charged membrane. **K-O**, 3D modelling of **F-J**. Image taken from[51].*



(a) **(b)**

Figure 1.7 *A comparison of cryo-EM images taken by Ryan Morris of lysate from samples containing (**a**) the supposed T4, confirmed to be T1 and (**b**) T4 obtained from the ATCC culture collection, thankfully confirmed to be T4. The difference in siphoviridic and myoviridic structures is clear.*

which have been categorized[139]. This under-characterization of protein function provides a small indication of the lack of focus on T1 in research. The phage has a reputation as a pestilence in most bio-labs, so much so that one of the most common *E. coli* mutations used in research is $\Delta TonA\Delta TonB$ which is immune to T1[53]. It is highly resistant to desiccation, can survive indefinitely as an aerosol, and completely immune to ethanol sterilization. It is also highly virulent and capable of out-competing every other T-series phage, as we discovered to our detriment. We initially believed our collaborator had sent us T4. However, after performing cryo-EM on some samples (figure 1.7) it became obvious that we were dealing with a different phages which was verified through short-read sequencing to be T1. The inclusion of T1 in this project was accidental and swiftly rectified with liberal application of a biocidal agent.

1.4 Osmotic Pressure and its Role in Phage and Bacteria

Key to the results presented in this thesis is an understanding of osmotic pressure and its role in bacterial growth and general physiology. Osmotic pressure arises

from differing levels of solute concentration across a semi-permeable membrane, wherein water will naturally flow towards the higher concentration - up the osmotic gradient. From the Encyclopedia Britannica, osmotic pressure was first properly studied in plants by Pfeffer in 1877. In 1886 van 't Hofft showed that if a solution is dilute enough that the partial vapour pressure above the solution follows Henry's law (that is, the weight of a gas dissolved in a liquid is proportional to the pressure of that gas on the liquid - i.e. that the partial vapour pressure is proportional to the solute concentration) then the osmotic pressure varies with temperature as if it were a gas occupying the same volume. Since the 19^{th} century the study of osmotic pressure has found relevance in a large number of areas in biology and colloid science[3][95][110][117].

This pressure arises from the lowering of the chemical potential of water via the dissolving of solutes. The chemical potential is itself a function of the free energy of water molecules, which can be directly linked to the kinetic activity of those molecules (also known as water activity). Solutes trap water molecules via charge interactions and thus reduce their kinetic activity (and thus water activity), which is a reduction of free energy. Since a region of low activity has many bound water molecules but few free water molecules it will draw in additional water molecules due to the increased free volume. Essentially, if there is space for molecules to move around they will tend to move into that space until the free volume is sufficiently reduced.

When two volumes of water are separated by a semi-permeable membrane, as in figure 1.8, water molecules will freely diffuse throughout each volume. If solutes are present in one volume but not the other, they will bind the water molecules, lowering the water activity and increasing the available free volume allowing more water molecules to enter that volume than leave. This causes a net flux inward to the solute-dense volume up until it is balanced by the pressure that volume puts on the semi-permeable membrane. This is a very physical description of the process, and a more thorough thermodynamic treatment is presented in Section 1.4.1.

In bacteria osmotic pressure has been studied quite thoroughly[3][62][84]. *E. coli* relies on a turgor pressure that arises from an osmotic gradient it maintains across its membrane to maintain its shape and grow larger. Water flows in to the cell in accordance with the higher internal concentration of solutes causing it to inflate. If the external concentration of solutes changes the cell will adapt by either raising or lowering its internal solute concentration to maintain a favourable

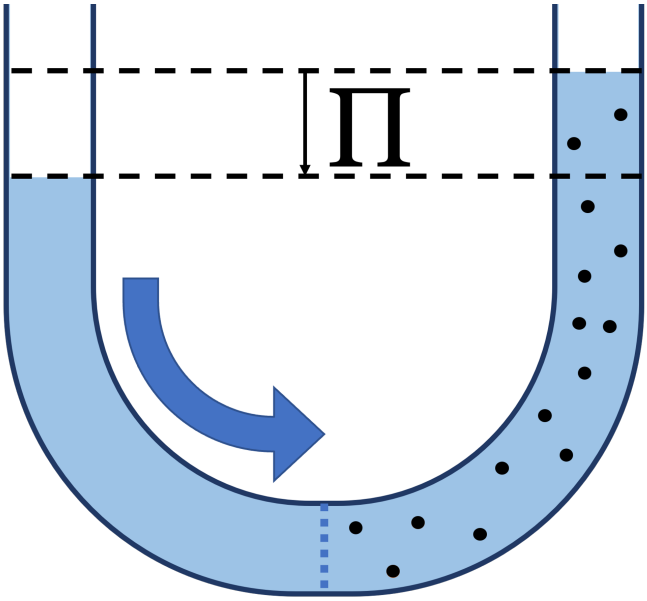


Figure 1.8 *Cartoon of water in a U-pipe. The dotted line in the center depicts a semi-permeable membrane through which water may pass but solutes (black dots) may not. Since the solute particles lower the water activity the number of interactions with the membrane is much higher on the pure water side. Thus, the flux of water will overall point from left to right and there will be more water in the right arm. As water flows into the right arm the pressure from the weight that resists the addition of new water molecules will increase. The osmotic pressure is the amount of this resisting pressure that brings the net flux of water across the membrane to zero.*

pressure of 0.3-3 atm[25][16]. The *E. coli* osmoregulation system is complex and involves different processes working in tandem to maintain turgor, an overview of this system can be seen in figure 1.9. *E. coli* undergoes morphological and physiological changes in response to osmotic shock. In response to hyperosmotic shock, where the external osmolarity is increased above the internal, the cell will shrink and plasmolyse (the inner membrane dissociates from the cell wall)[105] as water flows out of the cell in accordance with the now-reversed osmotic gradient. Following this the cell will begin importing K^+ ions and up-regulating pathways for organic osmolyte (solute) synthesis to increase its internal osmolarity such that the original osmotic gradient is maintained. In the case of hypoosmotic shock, the existing gradient is strengthened such that too much water flows into the cell causing it to distend. This results in mechanosensitive channels in the membrane opening allowing solutes to escape and causing the cell to shrink below its original size, after which it recovers through active osmolyte import as above[142][14].

1.4.1 The Thermodynamic Origin of Cellular Osmotic Pressure

Van 't Hofft's observation led him to conclude that one can consider the solutes as gas particles colliding with a membrane and it is that which produces the osmotic pressure. However this is slightly misleading as pointed out by Dick[27], as in the case of osmosis it is the solvent (water) that moves, not the solute. Additionally, this description does not adequately describe experimental data and though approximations are useful, there is a more thorough way of describing osmotic pressure through thermodynamics rather than simply kinetics.

Recalling from thermodynamics, the Gibbs Free Energy is written as

$$F = E + PV - TS \tag{1.1}$$

where F is the free energy, E is the internal energy, P, V, T are pressure, volume, and temperature, and S is the entropy. If the change in E is written as

$$dE = q - w, \tag{1.2}$$

or the difference in heat absorbed by the system and the work done, one can

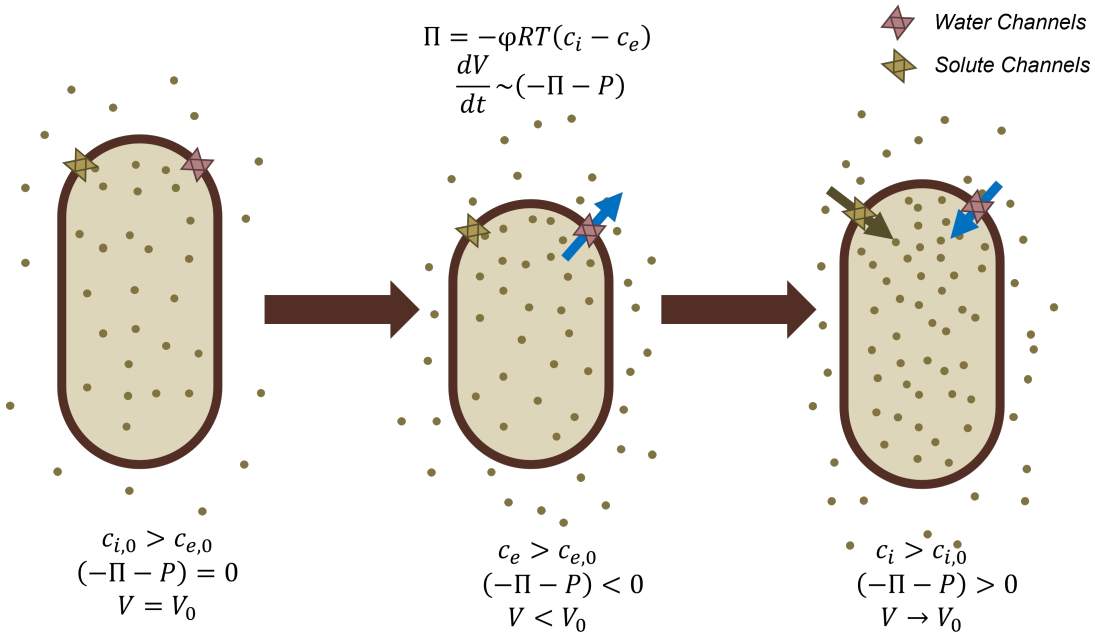


Figure 1.9 *A cartoon showing E. coli osmotic regulation during hyperosmotic shock. Above are equations for the osmotic pressure* Π *and the change in volume that is due to a balance between the osmotic pressure drawing water in and the resulting hydrostatic pressure inflating the cell. Solute channels are those that import or export solutes and water channels allow the passage of water either passively in the case of aquaporins or actively in the case of mechanosensitive channels. At rest a cell with positive turgor pressure is inflated by maintaining a higher internal concentration of solutes than the environment. Hyperosmotic shock occurs when the external solute concentration is increased causing an efflux of water as the osmotic gradient is decreased. The cell then actively increases its internal solute concentration to draw water back into the cell.*

express two relations for the volume and entropy of the system,

$$\left(\frac{\partial F}{\partial P}\right)_T = V, \left(\frac{\partial F}{\partial T}\right)_P = -S. \tag{1.3}$$

Water moves according to a chemical potential μ, and the chemical potential of a single component in a solution is its partial molar free energy, i.e.

$$\mu_1 = \left(\frac{\partial F}{\partial n_1}\right)_{T,P,n_2,n_3,\ldots}. \tag{1.4}$$

One obtains the effect of pressure on μ_1 by differentiating (1.3) with respect to n_1, i.e.

$$\frac{\partial}{\partial n_1}\left(\frac{\partial F}{\partial P}\right)_{T,n_2} = \frac{\partial}{\partial P}\left(\frac{\partial F}{\partial n_1}\right)_{T,n_2} = \left(\frac{\partial \mu_1}{\partial P}\right)_{T,n_2}, \tag{1.5}$$

and

$$\left(\frac{\partial V}{\partial n_1}\right)_{T,n_2} = \bar{V}_1, \tag{1.6}$$

such that

$$\left(\frac{\partial \mu_1}{\partial P}\right)_{T,n_1,n_2} = \bar{V}_1, \tag{1.7}$$

where $\bar{V}_1$ is the partial molar volume of component 1. For water this is usually about 18 mL. Applying an excess pressure, the osmotic pressure Π, will increase the chemical potential according to

$$\Delta\mu_1 = \int_{P_0}^{P_0+\Pi} \bar{V}_1 dP. \tag{1.8}$$

Since water is nearly incompressible the assumption that $\bar{V}_1$ is practically constant can be made, and the change in chemical potential becomes

$$\Delta\mu_1 = \Pi\bar{V}_1, \tag{1.9}$$

and thus the increase in the chemical potential of component 1 (the solvent, water) due to pressure is directly proportional to the increase in pressure with the coefficient of proportionality being the molar volume.

An addition of a second component (solute $\equiv$ osmolyte) will cause the chemical potential of the solvent to lower, by what amount one can calculate from the entropy of mixing the two substances. The entropy of a mixture arises from the probability of the components of a mixture to be in a highly ordered (unlikely), or highly disordered (likely) state, such that the *thermodynamic probability*, W, is

the number of molecular arrangements that are possible with a given macroscopic state. The bigger the macroscopic state, the higher the number of possible arrangements and thus the higher the W. For an individual molecule the entropy is given by

$$S = k_b \ln W, \tag{1.10}$$

where k_b is the Boltzmann constant, equal to R/N (where R is the molar gas constant and N is Avagadro's number). As the initial (pre-mixed) formulations are uniform in nature, the initial entropy S_0 must be zero, as there is only one way of arranging one thing and so $W_1 = W_2 = 1$. When the two components are mixed however the expression for $W_{1,2} = W_m$ becomes

$$W_m = \frac{(N_1 + N_2)!}{N_1! N_2!}, \tag{1.11}$$

and the entropy after mixing becomes

$$S_m = k_b \ln \frac{(N_1 + N_2)!}{N_1! N_2!}. \tag{1.12}$$

Provided n is large (which it is as N_1 and N_2 are very large) this may be simplified using $\ln(n!) = n \ln n - n$. Additionally N_1 and N_2 may be converted to gram-molecules by dividing by Avagadro's number, N. Thus the equation for the change in entropy becomes

$$\Delta S = R\left[n_1 \ln \frac{n_1 + n_2}{n_1} + n_2 \ln \frac{n_1 + n_2}{n_2}\right], \tag{1.13}$$

and the change in partial molar entropy of component 1 is obtained by differentiating by n_1,

$$\Delta \bar{S}_1 = \left(\frac{\partial \Delta S}{\partial n_1}\right)_{n_2} = -R \ln \frac{n_1}{n_1 + n_2}. \tag{1.14}$$

The change in free energy for component 1 is given by (after integrating for variation in n at constant temperature)

$$\Delta \bar{F}_1 = \Delta \bar{H}_1 - T \Delta \bar{S}_1, \tag{1.15}$$

where $\Delta \bar{H}_1$ is the change in partial molar heat content for component 1. It is assumed that $\Delta \bar{H}_1 = 0$, implying there is no temperature or volume change upon

mixing the two components. Thus

$$\Delta\mu_1 = \Delta\bar{F}_1 = -T\Delta\bar{S}_1 = RT\ln\frac{n_1}{n_1+n_2}, \tag{1.16}$$

where $\frac{n_2}{n_1+n_2}$ is a fraction of which the logarithm must be negative, and so the change in chemical potential must be negative on mixing component 1 with 2. One may use this relation to derive the classic expression (van 't Hofft) for osmotic pressure since at osmotic equilibrium

$$\Delta\mu_1(frompressure) + \Delta\mu_1(fromsolvent) = 0, \tag{1.17}$$

and so from (1.9) and (1.16) one may write

$$\Pi\bar{V}_1 + RT\ln\frac{n_1}{n_1+n_2} = 0. \tag{1.18}$$

If it is assumed that the solution is very dilute, then $\ln\frac{n_1}{n_1+n_2} \approx -\frac{n_2}{n_1}$, and the osmotic pressure may be written as

$$\Pi\bar{V}_1 = RT\frac{n_2}{n_1}, \tag{1.19}$$

and since $n_1\bar{V}_1 = V_1$,

$$\Pi = RT\frac{n_2}{V_1} = RTm, \tag{1.20}$$

where m is the volume-molar concentration which is essentially the molal concentration. This recovers the classic interpretation of osmotic pressure, but needs a lot of assumptions and approximations. It also implies that the solution in question is perfect (i.e. that the solutes are identical in mass, size, and intermolecular force) which is never true in biology. The point of explicitly writing out these derivations was to show the physical source and definition of osmotic pressure while also showing the necessity of a modification of the classic definition when applying it to living cells.

Since biological solutions are never really ideal, one may not expect to recover an accurate value for the osmotic pressure from (1.20). Instead there is a so-called empirical correction factor, or fudge-factor, ϕ, which is known as the osmotic coefficient and differs for each species of solute. With this factor the osmotic pressure can be written as

$$\Pi = \phi RTm, \tag{1.21}$$

or, considering a difference in solute concentrations across a membrane,

$$\Pi = -\phi RT(c_e - c_i), \tag{1.22}$$

where c_e and c_i are the concentrations of extracellular and intracellular solutes respectively. The negative sign modifies the pressure to act on the membrane pointing into the cell rather than pointing out as $c_e > c_i$.

1.4.2 Osmotic Pressure and Phage DNA Ejection

As mentioned in Section 1.2.2 there have been several models of DNA ejection that focus on osmotic pressure as the main driver of ejection both *in vitro* and *in vivo*. The idea that a phage capsid presents a high osmotic pressure arises from the fact that DNA is densely packaged inside the capsid and that both anions and cations are trapped in the tightly coiled DNA. It has been found that altering the external osmolarity by adding Na^+ one can decrease the spacing between DNA strands in a phage λ capsid, suggesting that bending energy cannot account for all of the energy landscape within a phage[107]. In the same work it was calculated that the energy in DNA self-interaction was as much as 38000 k_bT for a wild-type λ, an order of magnitude greater than that of DNA bending - thus DNA self-repulsion is the dominant contribution to capsid stress.

The hydrodynamic model proposes that the osmotic pressure inside a phage capsid can be calculated from osmolyte concentrations used in Evilevitch's osmotic suppression experiment. However, in the subsequent work by Jeembaeva, discussed in the DNA Condensation Model Section below, the result of full ejection depending on the presence of DNase rather than osmolarity somewhat casts doubt on the calculation of osmotic pressure through suppression. In terms of general osmotic study the phage capsid presents a difficult compartment for which to determine the osmotic pressure as the near-crystalline DNA does not allow enough water inside the capsid to qualify as a solution. Moreover, the DNA is just one molecule but its size means calculations regarding concentrations must be done by number of base pairs rather than number of DNA molecules and it is unclear what contribution if any a single base pair should have compared to its neighbours. Nevertheless, there does appear to be a flux of water and solutes from within the cell at the point of infection, and so the question of the relationship between DNA internalisation and osmotic regulation remains open.

There have been various attempts to provide a unifying theory for DNA internalisation, however there has yet to be a consensus. I here identify three separate hypotheses or models for this process: the internal pressure model (sometimes referred to as the continuum model or pressure conjecture) which was held in the nascent period of phage research[148]; the hydrodynamic model proposed by Molineux which describes water transport mediating DNA internalisation[92]; and a model of DNA condensation providing the necessary force for internalisation proposed by Jeembaeva [55].

The Continuum Mechanics Model

In the first and original hypothesis only the forces that arise from DNA being packed tightly are considered. DNA itself is a long molecular chain that is negatively charged, and so strong electrical repulsion is generated when it is forced together without the aid of positive charges like histones (in the formation of chromosomes[5]). In a phage capsid DNA is packaged up to a maximum concentration of 500 mg/ml[120] resulting in a pressure of around 20 atmospheres. It was shown in 1968 that by packaging nylon wire into a model glass phage to a pressure of 30 atm it resulted in ejection of the wire at phage timescales, however it is unclear that frictional forces would be conserved at different scales[148]. Despite this, this result was used as evidence in support of an internal DNA pressure being the driver for the ejection process - the difficulty of obtaining meaningful data prior to single molecule studies has been cited as a reason for the lack of greater insight [101]. In phage λ DNA ejection has been studied well *in vitro* and the results seem to accord with both the nylon experiment and the continuum mechanics model as a whole. However, as shown by Leforestier, phage T5 demonstrates a different behaviour as it will not eject the same length of DNA for all pertinent environmental conditions (i.e. what it may expect to encounter in nature) which suggests that this does not hold true for all phages and thus there must be additional un-resolved factors[72].

The Hydrodynamic Model

This second hypothesis attempts to bridge the gap between these different ejection mechanisms by suggesting an alternative - species neutral - mechanism. The logic behind this model is that the capsid-contained DNA traps both anions and cations

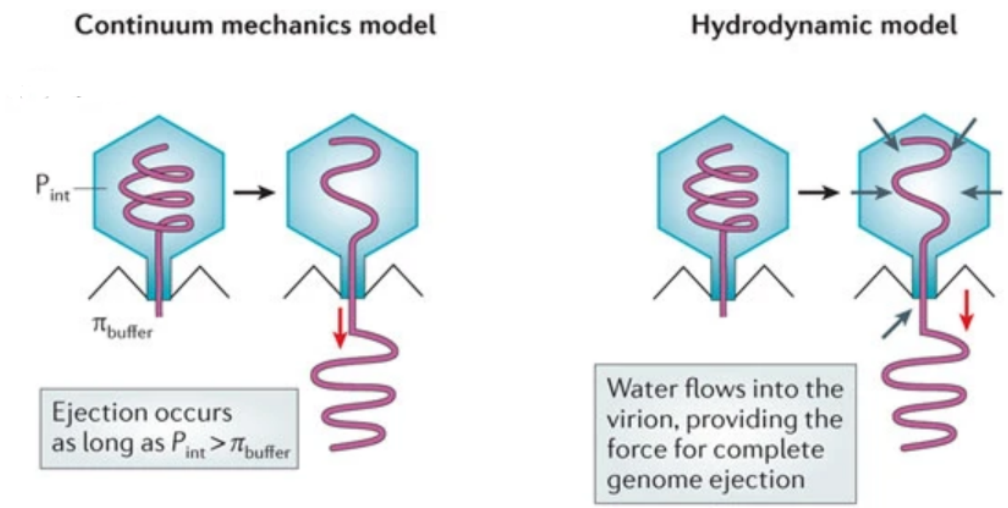


Figure 1.10 *A graphical comparison of the continuum mechanics and hydrodynamic models adapted from[92]. P_{int} is the internal pressure arising from DNA-DNA repulsive forces and bending forces, and π_{buffer} is the osmotic pressure of the environment.*

within its tightly packed conformation (as seen in fig 1.11)[75], causing a large apparent ionic density and thus a strong osmotic gradient towards the center of the capsid. Since the DNA is tightly packaged and the capsid is closed, there is no opportunity for the osmotic gradient to balance and the pressure remains. When the tail or capsid opens the resulting hole allows the DNA to escape, water will flow into the capsid through the protein shell and out through the hole, bringing the phage DNA with it.

This model *could* hold *in vivo* as the opening of a hole in the side of a bacterium will cause the cellular pressure to collapse (as shown in Chapter 4), leading to first a short efflux and then a long influx of water. It is possible that the influx of water could drag the DNA into the cell, however it is not clear if this agrees with experimentally observed timings. There is currently no experimental evidence that would suggest that this is the case, and even *in vitro* it does not have a particularly strong argument. In terms of solution the interior of the phage capsid is highly non-ideal (ϕ is either much higher or much lower than 1) and so the osmotic pressure is very difficult to define. It is not clear that, if the phage capsid should open, water should continually flow into it and propel the DNA out - especially as the model assumes water flows *out* of the tail pulling DNA with it.

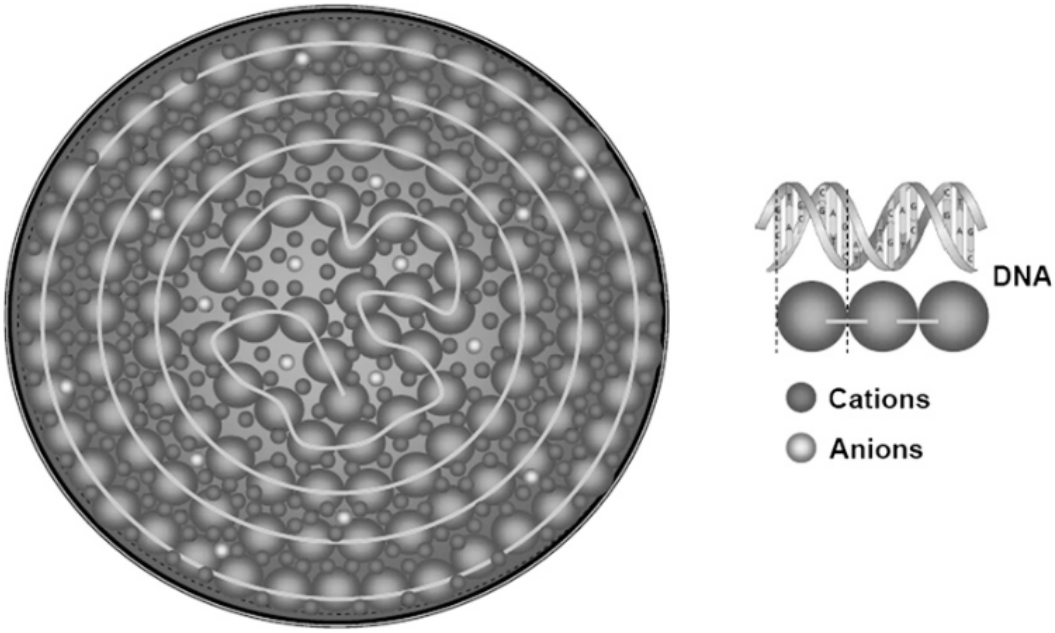


Figure 1.11 *Cartoon of DNA packaged within a phage capsid. Large circles connected by a grey line represent the DNA chain packed into the capsid. Near the centre of the capsid the organised packing breaks down and the DNA molecule must be bent beyond its persistence length[92], and the black and white dots are the respective cations and anions trapped by the DNA molecule. Image taken from[75].*

DNA Condensation

The previous model by which osmotic pressure is the primary driver for DNA internalisation cites as supporting evidence work by Evilevitch termed the 'osmotic suppression experiment'[32]. In this work it was demonstrated using phage λ that a sufficiently high concentration of PEG8000 would raise the external osmolarity so high it would suppress DNA ejection when λ receptor binding proteins were present in solution. As the external osmotic pressure was varied from 0 to 20 atm, the amount of ejected DNA decreased from all to none. This work was performed with a nuclease (DNase I) present in solution, which breaks down DNA by cleaving at particular sites leading to a solution of free nucleotides. Since the capsid is impearmeable to nuclease only the DNA that was ejected was broken down and so can be measured by spinning down the mixture after the ejection has finished.

However, work performed later in the same group by Jeembaeva[55] suggests that the experiment may not support the Hydrodynamic model as is assumed. It was found that in the absence of nuclease one observes the ejection of the full genome even at high external osmolarities. The the DNA Condensation Model (figure 1.12) was proposed, providing the argument that it is precisely *because* the external volume has a high osmolarity that the DNA can fully eject.

The DNA condensation model predicts that a complete genome is internalised by molecular-crowding-induced DNA condensation. A DNA molecule in a crowded environment will reduce its volume to maximise molecule-solvent interactions. This reduction in volume will cause depletion reactions within the resulting globule giving rise to a strong an-isotropic osmotic pressure acting towards the depleted regions (de-hydrated regions) in the DNA. It is posed that this condensation of the DNA provides the complete force required for the entire genome to be internalised, and does not depend on any existing pressure gradients across the membrane - only requiring the cytoplasm to be crowded with macromolecules.

Previously it was shown that non-specific DNA binding proteins HU and DNase I can speed up the ejection process *in vitro* by exerting 'pulling forces' on the molecule and acting as molecular motors[80]. The argument presented in the DNA condensation model is that despite the earlier result that high osmolarity suppresses ejection, the molecular crowding within a cell cytoplasm actually *promotes* ejection of a *complete* genome through crowding-induced DNA

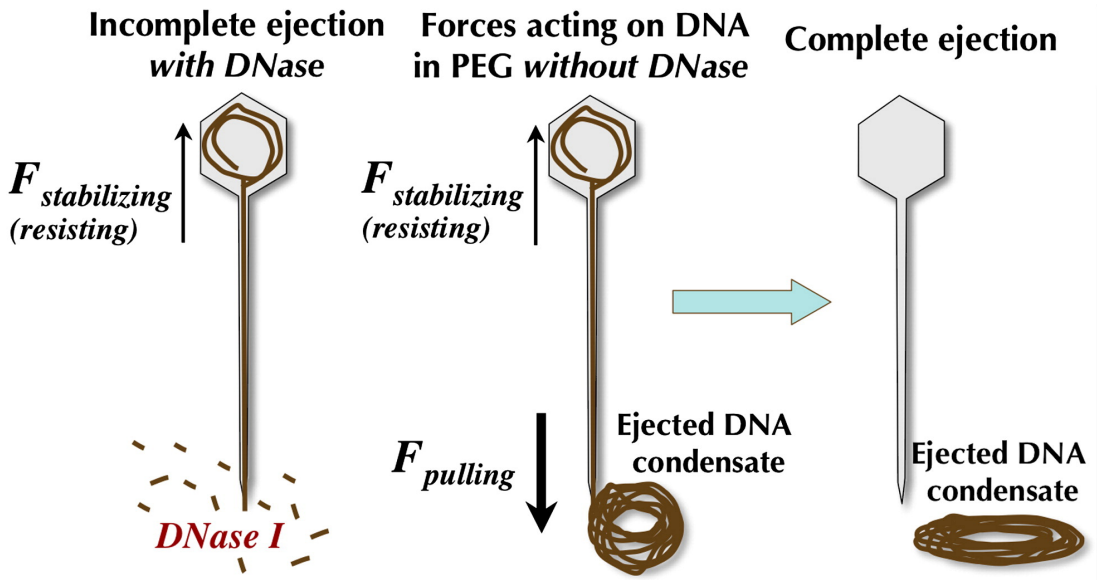


Figure 1.12 *A cartoon depicting the forces suggested by the DNA condensation model. While DNase I is present there can be no DNA condensate formed and so the only force present acts to stabilise the DNA in the capsid. When DNase I is absent both capsid and ejected DNA condensates are subject to a force arising from the same interaction. When the environmental osmolarity is raised (in vitro by adding PEG8000), the pulling force dominates and DNA is ejected. Image taken from[55]*

condensation (condensation here referring to the adoption of a concentrated, compact state occupying a small amount of the available volume). Molecular crowding in an environment will lead to steric competition among the solutes, changing the water activity and dehydrating DNA. Such excluded volume effects have long been suggested to be important for DNA condensation as they favour compact molecular conformations[73]. It has been suggested with experimental support that in cytoplasm the condensation of DNA into so-called nucleoids is driven by a combination of these 'direct' crowding effects and 'indirect' effects such as the increased binding of histone-like proteins under crowded conditions[149].

Concluding Remarks on the Three Models

While some specific phage may have well-defined DNA injection processes, it should be clear by now that there is no general consensus on an over-arching injection mechanism. This thesis - with the exception of Chapter MODEL - focuses largely on the first step of DNA injection, that of the formation of a bridge from phage capsid to cytoplasm. It is broadly an attempt to understand the physical consequences of the mechanical or enzymatic creation of a hole from the extracellular to the intracellular. The models of injection presented above each colour any possible explanation of the phenomenon highlighted in this thesis, so it is important to evaluate them as to their likelihoods.

Firstly, the Continuum Mechanics model provides a simplistic approach to the problem that can in theory explain simple phage ejections (such as T4's one-step process). However given the wide variety of ejection mechanisms it seems that it is unlikely that this model can explain all. After all, if a phage could rely on simple bending pressure to inject its DNA, what is the use of using complex protein machinery to do so instead? The Hydrodynamics Model seeks to explain this with some support experimentally, that a high-osmolarity environment does not support the total ejection of the genome. However this model has no experimental evidence to support it and ultimately does not account for subsequent experimental results showing full ejection in a high-osmolarity environment without the use of DNase. For phages that perform their ejection in a single step, it seems that the condensation of DNA as presented in the DNA Condensation Model, has the best experimental evidence supporting it. However again this poses the question, if phages could rely on a physical process to inject

DNA why do some evolve such complex mechanisms? This is a question that has yet to be sufficiently answered.

Chapter 2

General Methods and Materials

The methods used in this thesis fall into two broad categories. The general methods presented here have been created by others and are used throughout the thesis. Bespoke methods developed by me are presented in Chapter 3. In both cases I present them as protocols for ease of repetition.

2.1 General Methods

2.1.1 Strain Information and Stock Upkeep

The strain of *E. coli* used was EK03, a mutant strain of K12 MG1655 with *fliC* replaced by $fliC^{sticky}$ on the chromosome and a pkk223-3/pHlourin plasmid containing *pHluorin* gene under an IPTG-inducible *Tac* promoter. This gene is GFP/YFP fluorescent and was originally used for measuring changes in *E. coli* cytoplasmic pH, but is also suitable for general cell size experiments[104][105]. The strain was produced by Dr E. Krasnopeeva prior to my arrival to the group. The *fliC* replacement was performed for tethered-cell and bead-assay PMF measurements by allowing either the cell to be tethered to a cover slip by the sticky flagella, or a bead to be stuck to the flagella as it rotates. Initially there were some plans to perform PMF measurements alongside the volume size measurements during phage infection, hence why this strain was used rather than a simple K12 strain expressing a fluorescent protein only.

The strains of phage used were T1, T4, and T7. T4 and T7 were procured from the

DSMC (reference number DSM 4623) and ATCC (reference number 11303-B4) culture collections respectively. T1 was provided accidentally by our collaborator instead of T4. As T1 is highly likely to become a difficult contaminant we disposed of it once identified. I continued with solely T7 after seeing reproducible infection with shortness of time preventing expansion of experiments into other phage.

The growth medium used for all experiments was Neidhardt's EZ Rich Defined Medium (RDM) [97]. It is a defined medium that results in consistent growth rates close to those found in the undefined Luria broth (LB), the most common bacterial growth medium. We prefer a defined medium over an undefined medium as it is a) easy to modify by omission or addition of desired components and b) the growth of bacteria is more consistent. In LB there are unknown quantities of amino acids acting as the carbon source. As *E. coli* grows it will preferentially switch between amino acids as they deplete, from the acid providing the highest growth rate to the acid providing the lowest in a process called catabolite repression[39]. This switching process may disturb the growth so that on a per-bacterium basis the growth at a given point in time is not necessarily exponential. In RDM there is generally only one carbon source - in my case glucose, and this source is added in saturation so the bacteria cannot deplete it prior to reaching stationary phase. Our RDM was created from frozen stocks prepared by our lab technician, Tracy Scott.

In RDM with glucose (RDMG), EK03 has a doubling time of approximately 24 minutes at 37°C. I maintained a frozen stock at -80°C of small aliquots of EK03 in a 2:3 RDMG:50% glycerol solution. The frozen stock was generated with the following standard protocol from the Pilizota Lab:

Frozen Stock Preparation

1. Spread a loopful of reference stock of EK03 on an LB-agar plate with 10 mM ampicillin.
2. Grow plates overnight at 37°C, collect early in the morning for small colonies.
3. Select a single colony and add to 10 mL LB plus 10 mM ampicillin in a conical flask and grow overnight.
4. In a small falcon add 3 mL overnight culture and 2 mL 50% Glycerol solution for a 3:2 ratio of LB-culture:Glycerol.

5. Portion LB-culture:Glycerol solution into Eppendorf tubes (100-130 μL each).

6. Store at -80°C.

I generated a growth curve for EK03 using a 20×10^4 times dilution into 10 mL total volume fresh RDMG. Bacterial growth for each experiment was performed in small incubators located in a 4°C cold room on a timed circuit. The culture would be stored overnight in the incubators which would turn on approximately 6 hours before my experiments began to ensure a consistent OD_{600} for each experiment.

Preparation of stocks of T7 was performed using the following simple amplification protocol, with the first propagation from T7 obtained from the DSMZ collection as reference:

Phage Stock Preparation

1. Inoculate EK03 in 10 mL total volume RDMG in a sterile conical flask. Incubate at 37°C overnight with 200 rpm rotation.

2. Retrieve flask in morning and centrifuge at 4000 g for 5 min at 4°C. The OD_{600} value should be 2-3 for a cell count of approximately 10^9 cells/mL.

3. Re-suspend the pellet in fresh RDMG to a total volume of 10 mL.

4. Add 20 μL of reference T7 phage stock to flask and incubate at 37°C with 200 rpm rotation, checking every 30 minutes for clearance.

5. Once cleared, retrieve flask and centrifuge at 4000 g for 5 min at 4°C.

6. Filter supernatant through a 0.22 μm syringe filter and store at 4°C.

I kept a reference stock so that each new stock of T7 I produced would still be only one generation removed from the reference stock. Phage titre was determined using the double agar overlay plaque assay technique [67]:

Double Agar Overlay Plaque Assay Technique

For the hard agar plates:

1. Prepare agarose medium to 21 g/L and heat sterilize.

2. Prepare agar plates with 18-25 mL of medium per plate. Store at 4°C once cool.

For the soft agar overlay:

1. Prepare overlay medium to 8 g/L agarose and 13 g/L nutrient broth and heat sterilize.
2. Divide into aliquots of 3 mL and hold at 55°C.

Preparation of bacteria:

1. Prepare overnight culture of EK03 cells in RDMG the night before the plaque assay. Collect in stationary phase.

Preparation of phage:

1. From T7 stock flask dilute 0.1 mL into 0.9 mL of RDMG in an Eppendorf tube.
2. Serially dilute into further Eppendorf tubes 0.1 mL at a time until the desired range of concentrations is achieved (typically 10^{-8}-10^{-11}).

Double agar overlay counting of phage:

1. Remove number of prepared agar plates from storage such that 3 plates can be used per phage dilution plus 3 plates for a phage-free control.
2. Dry agar plates to remove condensation.
3. For each dilution add 100 μL of phage dilution and 100 μL of EK03 culture to one soft agar aliquot and shake gently.
4. Pour soft agar aliquot over the hard agar plate, ensuring even distribution.
5. Allow to harden, then incubate at 37°C for 3-5 hours, checking regularly for plaques (visible holes in cloudy agar overlay).
6. Count plaques on plates containing 50-300 phages.

Stocks would be refreshed when they fell below 1 mL and the protocol typically produced approximately 10 mL of clear phage stock at a titre of roughly 10^{12} pfu/mL. Pfu/mL refers to plaque-forming units per millilitre. A plaque is the result of an infection by one phage in the original mixture, and so the number of plaques per mL gives a measure of the number of phage in solution at time of addition to bacteria. After about a week one expects the titre of phage to decrease by a factor of ten before the concentration stabilises (based on experience), and so the long-term concentration of my phage stocks was about 10^{11} pfu/mL. Phage titre was determined after the first couple of stock refreshes to make sure that the protocol was consistent. After consistency was reached, the titre no longer needed to be constantly checked however was still occasionally monitored to make sure it was not drifting out of the expected range.

2.1.2 Home-Made Microfluidic Devices

The microfluidic chips used in the cell sizing experiments are simple in-house-built devices consisting of a glass cover slip and microscope slide held together with double-sided scotch tape. There are two designs, illustrated in figure 2.1. The Tunnel Slide was used for the initial qualitative experiments in killing within a microfluidic chamber as the simplicity allows many repeats to be made within a short period of time. The Flow Cell was used for quantitative measurements as the attached tubing allowed the growth state of the bacteria to be maintained with fresh media.

Flow Cells were created and stored as 'blanks' which did not have a chamber. Blanks were created using the following protocol:

Flow Cell 'Blank' Creation

1. Draw the required chamber dimensions (2x40 mm) on a glass slide, marking positions for inlet and outlet at either end of the chamber.

2. Using this slide as a template, mark inlets and outlets on fresh slides.

3. Build a small well using sticky tack or equivalent on each slide around each mark and fill with tap water. Ensure no leaks as this keeps the glass cool and prevents fractures from forming.

4. Using a 1 mm engraving drill (in this case from Dremel) in a small drill

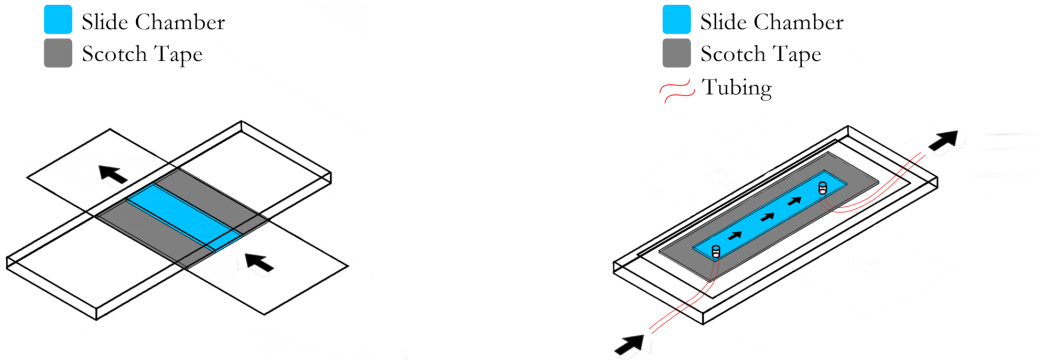


Figure 2.1 *Cartoon comparison of the Tunnel Slide (left) and Flow Cell (right). The Tunnel Slide is a simple device with a chamber that is filled by capillary action. Solution is wicked through the chamber using tissue and the slide is sealed prior to imaging. The Flow Cell is connected to tubing allowing the continuous flushing of solution through the chamber. The sketched devices are inverted with respect to each other to best show the chamber and liquid flow. In both cases the chamber lies on the bottom (objective side) of the chip.*

press gently drill the inlets and outlets.

5. Repeat Steps 3 and 4 until all slides are drilled.
6. Clean all slides using liberal amounts of tap water.
7. Use two-part epoxy to glue 0.07" OD tubing into the drilled holes and leave to set overnight. Ensure the tubing is poked through in the opposite way to drilling as this means the chamber-side of the hole is cleaner-cut.
8. When epoxy is dry trim tubing on the underside of the chamber using a flat razor blade and store under cover.

Typically the input tube was kept shorter than the output tube, as the chambers were loaded with small amounts of fluid and so input tubing volume was minimized. As such, the amount of cell and phage stocks and bead solutions used per day were kept low, reducing the pace at which stocks were depleted.

Extension tubes were used to connect the Flow Cell to the syringe pump and the waste falcon. They served a dual purpose in that they allowed the blanks to have shorter tubing permanently fixed, easing storage, and they gave the tubing in the system a degree of freedom in the length. A short permanent inlet was useful as the loading volumes were small, in general its necessity depends on the volumes used. They were made with the following protocol:

Extension Tubing Creation

1. Cut appropriate lengths of 0.07" OD tubing. Ensure enough is available to reach the syringe pump and the waste from the stage.
2. Using a small pair of snips, cut the metal part of a brown gauge syringe off and cut it in two, dispose of the plastic part.
3. Using a pair of pliers, squeeze the snipped ends back into an open shape.
4. Insert one half-piece into one piece of tubing such that half the metal piece is sticking out acting as a connector to the Flow Cell tubing.

Completed Flow Cells were created each morning before the day's experiments began according to the following protocol:

Flow Cell Construction

1. Clean one blank with ethanol and distilled water.
2. On reverse side of template place double-sided sticky tape lengthwise, completely covering the chamber.
3. Using a razor blade cut out the chamber using the template as a guide. Avoid over cutting on corners as this will induce leakage.
4. Remove sticky tape from template, leaving the cut-out portion on the glass.
5. Turn template over and place a 24x60 mm glass cover slip over the drawn chamber.
6. Place the sticky tape onto the cover slip gently over the drawn chamber. Use a **clean** pipette tip to gently tap the tape down onto the slip.
7. Gently place cover slip and tape onto clean master.
8. With a pipette tip push on the cover slip above the tape to seal it properly. Light-coloured areas of tape are more prone to leakage.
9. Repeat for the desired number of complete Flow Cells.

After each experiment the Flow Cells were deconstructed and cleaned so they could be stored as blanks for the next set of experiments. The cleaning procedure is as follows:

Flow Cell Cleaning

1. Using a syringe, flush 3-5 mL of 70% ethanol solution through Flow Cell chamber. Use force such that the cover slip flexes and the tape begins to separate to make Step 3 easier.
2. Flush 3-5 mL of de-ionized water through the chamber. If leakage occurred during ethanol flush use pressure to hold chamber closed.
3. Using a clean razor blade, **carefully** cut away the cover slip. If the cover slip begins to break up and is very difficult to remove, simply dispose of the Flow Cell to avoid unnecessary risk.
4. Using a tissue soaked in ethanol remove remaining tape residue and wipe slide dry and store under cover.

Chapter 3

Methods Developed for This Thesis

In addition to the General Methods used in this thesis are a set of bespoke methods developed by me to produce and analyse the experimental data for the results presented in Chapter 5. This chapter is split into three parts. The first two comprise an experimental method for the gathering of video data of single *E. coli* cells for analysis, and the subsequent development of an analysis method that would culminate in the use of an existing neural-network pipeline. In each case there are comments on the methods and the decisions made during their development. The third part is somewhat monolithic as it describes work undertaken at the beginning of the project to adapt an in-house-built microscope for the measurements.

3.1 Cell Size Measurements and Data Analysis

3.1.1 A Protocol for Measuring Cell Size in a Flow Cell

Cell sizing was carried out in Flow Cells using a protocol initially adapted from Prof. C. J. Lo, who provided preliminary results showing the volume loss (as in the Lay Summary, termed 'dip'). This preliminary protocol had several caveats, namely in that it used a Tunnel Slide (figure 2.1, left) and so was a closed system. Due to this, the bacteria would not grow normally as they would deplete their carbon source and so one could say very little about their physiological state prior to infection. The decision was made to create our measurement protocol in

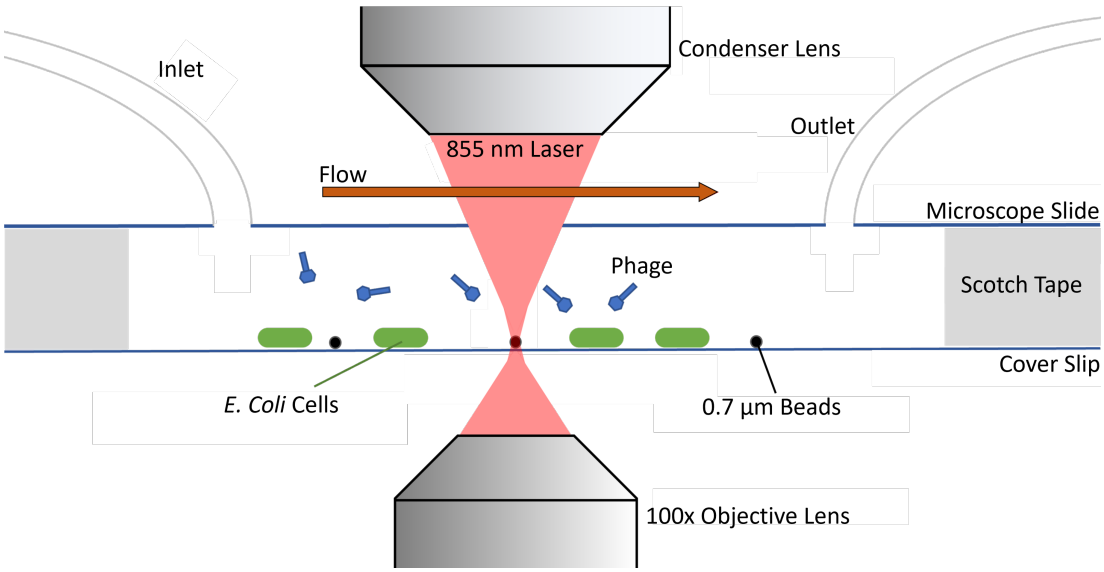


Figure 3.1 *Cartoon cross-section of a Flow Cell during the cell-sizing measurement. Shown are the E. coli cells and 0.7 µm beads adhered to a thin layer of poly-L-lysine, over which fresh media containing phage is flowed. The 855 nm laser (bottom to top) is focused on a single bead allowing the stage to be held still in XYZ through a back-focal-plane interferometric feedback loop (discussed in Section 3.3). Inlet is connected to a syringe pump and the outlet is connected to a waste Falcon tube. Cartoon is not to scale.*

a Flow Cell (figure 2.1, right) so that media could continuously flow over the cells such that they would not deplete any nutrient. A schematic of the experimental setup can be seen in figure 3.1. The finished protocol developed to measure the size of single phage-infected bacteria, the results of which are found in Chapter 5, is:

Single-Cell Size Measurement

1. The day before, dilute one aliquot of frozen stock of EK03 20,000x into RDMG and incubate to an OD_{600} of 0.15.

2. In the morning, prior to collection of EK03, make 1mL of 10^{10} pfu/mL dilution of phage.

3. Make up the required number of Flow Cells, usually 2 or 3 depending on the length of the day. Ensure the inlet pipe is short and the outlet is long.

4. If necessary, make enough extension tubing for the Flow Cell by cutting a syringe needle in half and attaching to sufficient lengths of tubing.

5. Make 1 mL of 100x dilution of 0.7 μm beads in RDMG. Wash into fresh RDMG twice using the bench top centrifuge at 17,000 g for 5 minutes each time.

6. Prepare a flask of 9.7 mL RDMG.

7. Collect overnight culture and inoculate 300 μL into the prepared RDMG flask. Incubate at 37°C for approximately 2 hours with 200 rpm rotation.

8. Make 3 1 mL aliquots of overnight culture and centrifuge at 17,000 g for 5 minutes.

9. Discard 900 μL of supernatant from each aliquot.

10. Re-suspend remaining culture into one tube and add 100 μL of fresh RDMG for a total volume of 400 μL.

11. Flush poly-L-lysine (PLL) into the Flow Cell through the output (long) pipe, ensuring the whole chamber is filled. Stop flushing when the PLL reaches the inlet hole.

12. Attach syringe containing 1 mL of RDM to inlet tube, detach PLL syringe and attach extension to outlet, ensuring end of extension is in a waste falcon tube.

13. Flush 1 mL RDMG through the chamber. Pulse syringe with hand to remove bubbles.

14. *When detaching syringes ensure a droplet of liquid remains in needle, when attaching make sure a bead of liquid on the syringe head couples with this bead in the needle to ensure no bubbles are formed. Before detaching syringe use forceps to clamp outlet shut.*

15. Flush 400 μL of culture into chamber, incubate for 5 minutes.

16. Flush 1 mL of RDMG through chamber, pulsing lightly to flatten cells.

17. Flush 200 μL of RDM-bead solution into the chamber, incubate for 5 minutes.

18. Refill RDMG syringe with 1 mL of RDMG and connect a fresh needle. Attach extension tubing to needle.

19. Detach RDMG-Bead syringe and needle from inlet tube, leaving small bead of liquid on tube end. *This step is highly delicate and must be done carefully.*

20. Attach extension tubing to inlet tube by coupling water droplets.

21. Flush 1 mL of RDMG through chamber.

22. Refill RDMG syringe and attach, do not flush yet.

23. Fill a fresh syringe with 1 mL of phage dilution and move slide to microscope.

24. Under bright-field find Köhler illumination using field stop aperture. Stop when aperture is in focus.

25. Find a FOV that contains many flat cells.

26. Switch imaging path to EMCCD (main imaging camera) and set excitation and emission filters to YFP.

27. With any stage stabilisation **off**, flush 1 mL of RDMG through the chamber and attach the phage syringe to the inlet. This refreshes the nutrients in the chamber.

28. With any stage stabilisation **on**, image for 5 min at 0.5s exposure with 10 s between frames (10 s frame-time).

29. With imaging **paused** and stabilisation **off**, switch to bright-field (adjusting gain and exposure time).

30. Hand flush approximately 200 μL phage into chamber. Place syringe in syringe pump and set to 0.1 μL/min. Start pump..

31. With imaging **resumed** and stabilisation **on**, image for 15 min with a 5 s frame-time (120 frames).

32. At frame 120, set frame-time to 10 s and image for 1-1.5 hours or until the FOV is blank.

33. Clean previous slide and prepare next experiment from Step 8 with the culture from step 7.

There are a few points of discussion regarding this protocol. To begin, as seen in Step **11**, this protocol uses thin layers of PLL to stick cells to the cover slip. The antimicrobial properties of PLL are well known and various authors have raised concerns about its use in such conditions. However, these concerns are valid for free-floating PLL molecules only, and work done previously in the Pilizota group shows that this method of cell attachment does not significantly impact the cells compared to other methods [136]. It is not expected, given this work, that the PLL coating will affect the membrane potential of the cells. If, however, there is a depolarizing effect this will inhibit the initial opening of a hole by the phage - thus for an experimenter looking to characterise this process examining different methods of sticking bacteria to the cover slip would be an ideal start [60].

In terms of flushing, those performed up to Step **31** were done by hand at rates between 5 and 10 mL/min. The high flow rate means the shear forces on the bacteria were quite strong which necessitated the initial concentration step. This shearing stripped many bacteria from the slide but had the benefit of flattening those that stuck. This significantly increased analysis yield despite lowering overall cell density. There was a gap of approximately one minute at step **30** due to the changing of syringes. The stage was stabilised through back-focal-plane interferometry (discussed in Section 3.3) which could not hold while syringes were swapped due to sudden pressure changes in the Flow Cell chamber. The stage must be stabilised during filming, and so for this step imaging was paused. In principle this could have been avoided by using a valve, but as the gap did not occlude the area of interest in the data it was not seen as a priority to adapt to. Finally, the change of frame-time to 5 s after adding phage was an attempt to catch shorter dips by increasing time resolution by 2x. Initially it had been set at a 2 s frame-time, but this had an adverse effect of cumulative cell damage related to exposure of the cell to low wavelengths of light [65]. Using a frame-time of 5 s allowed images to be captured at an adequate time resolution without compromising bacterial health.

Additionally, two data sets have been generated by this protocol. Lo's experiments had all been carried out at room temperature, and so initially the preparation (Steps **8** to **22**) and filming was done at about 20°C. During the process it was decided to film at 37°C as it was found that cells were not increasing in size at the start of the video - hence despite the continuous refresh of media it was still difficult to make any judgements about their growth state. The preparation, however, was still at room temperature. Several experiments

were done with these conditions before the decision to switch to preparing at 37°C in a hot-room was made. Thus, there is a set of data gathered with a 20°C preparation (dubbed Cold Shock) and a set of data gathered with a 37°C preparation (dubbed No Shock). Both sets were filmed at 37°C.

Images were taken on an inverted oil-immersion microscope constructed in-house. The fluorescence excitation light was generated by a white LED at 0.5 V shone through a 500 nm YFP (yellow-fluorescent protein) filter. Work by various ([65] and unpublished) in the Pilizota group has shown that photo-damage to cells does occur at this level however is minimized by short exposure times and infrequent measurements. Therefore the initial frame time was 5 seconds with 0.5 second exposure times for the first 15 minutes where the dip was thought most likely to occur, then after 15 minutes the frame time increased to 10 seconds to aid bacterial health. Over an hour this corresponds to approximately 450 images, or 225 seconds of exposure time. The photo-damage is cumulative and results in cell death at approximately 12 minutes at 0.65 V (unpublished), and so the regime used for these measurements should not result in cell death due to photo-damage.

3.1.2 Experimental Difficulties Overcome During Protocol Development

The initial step to develop a protocol of observing the interaction between bacteria and phage is necessarily to observe successful infections. This is trivial in bulk, wherein one must simply add phage to a growing culture of bacteria and wait for the solution to change from cloudy to clear. Observing this on the single-cell level is, predictably, more complicated and involved some problem solving, especially regarding flow control in the microfluidic chamber. This section describes problems encountered and their solutions, and some ideas for future improvements that are made more concrete in Section 3.4.

Flow Versus Static Conditions

The initial challenge to solve had been to improve bacterial growth rates and patterns to roughly match bulk conditions as it is difficult to make even a qualitative statement of any physiological effect if the bacteria is not growing well. This challenge was solved by changing the experimental conditions so that

the bacteria were imaged at 37°C and had a continuous flow of media refreshing them(by virtue of a Flow Cell). However, these changes brought unexpected challenges. Initial experiments indicated a difference in how phage infect in Flow Cells and Tunnel Slides. Experiments in the former would show significantly reduced killing compared with the latter, despite comparative concentrations of bacteria and phage in each. In both devices the number of bacteria in a given FOV were approximately equal, in the Tunnel Slide around 80-90% of the bacteria in the FOV die (disappear) and in the Flow Cells only 20-40% of bacteria would typically die.

In general I was able to consistently obtain 40-60 bacteria per FOV, which corresponded to approximately 10^8 cells/mL. The inspiring work by Lo had MOIs of 100 (i.e. a ratio of 100 phage to 1 bacteria), justified as in his Tunnel Slides, the ideal MOI of 1:1 was not sufficient for complete killing across the FOV. However, the challenge of reduced killing in Flow Cells remained present even at MOIs of 1000, indicating that it was not simply the amount of phage present that caused this problem.

Recalling to Section 1.3.1, the T7 bacteriophage diffuses across the bacterial membrane by attaching and detaching - effectively hopping across the surface. In order for it to bind and infect the phage needs to land 'on all fibres'. While the cause of non-infection was not clear at the time of experimentation, it seems very likely that the root cause is that the flow across the membrane in a Flow Cell was enough to disrupt the binding process. This is supported by the fact that when the flow rate was significantly reduced, to only a few millilitres per minute, the amount of observed killing increased to nearly 100%. Additional experiments comparing flow rate to FOV clearance could be helpful in understanding this link.

Another challenge introduced by flow conditions are air bubbles. These pose a significant issue in imaging as if they pass through a FOV they will inevitably disrupt stage stabilisation and the time-lapse will be ruined. Additionally it appeared that the introduction of bubbles to the chamber prior to going under the microscope caused a reduction in killing. It is not currently known why this was the case as these kinds of errors naturally decrease over time, but it may be that they were disrupting the membrane of the bacteria in some way or interacting with the thin film of PLL on the cover slip. To qualify the overcoming of this challenge, air bubbles are introduced into the Flow Cell due to error in experimentation, wherein hands moving too quickly or syringes pulled from tubes without care allow them to be introduced. Sadly, the only way this problem could

be resolved was to improve experimental technique. A good piece of protocol development for future work then would be to investigate ways of making the procedure robust to small mistakes.

Avoidance of False Positives Due to Cover slip Flexion

A challenge linked to the large flow rates above, which has prevented any data from T1 from being presented as there were no T1 experiments performed after this challenge was discovered, was that of false volume dips caused by pressure inside the Flow Cell chamber. When a syringe pump is activated the impulse causes an instantaneous pressure inside the microfluidic. Since water is effectively incompressible and thin glass is flexible this would cause the cover slip on which the bacteria are adsorbed to bulge outwards slightly. As this pushed the bacteria slightly out of focus and the stabilisation at the time was not quite able to adapt quickly enough it appeared on the volume traces as a dip. Pausing recording between Step **29** and Step **31** above prevented these false positives from being shown. The real dips were not expected to fall within this time-period, so no measurements were expected to be missed due to this pausing. It would be beneficial to examine whether or not this assumption is true, as it could be artificially decreasing the number of observed dips and could shed more light on the adsorption rates of different phage.

3.2 Extracting Single-Cell Area Data From Videos

Analysis of the videos was initially performed by a pair of LabVIEW programs written by Prof. T. Pilizota to measure the area of single bacteria by counting the number of pixels above a given threshold (figure 3.2). After finding that this technique was not quite fit for purpose due to being unable to analyse cells in contact with each other or sitting very close together I first attempted to write my own edge-finding protocol (figure 3.3) before settling on a machine-learning technique from the Dunlop lab: DeLTA2 (figure 3.5). DeLTA2 allowed me to extract area data from videos of single cells that were quite crowded, which both previous techniques struggled with. This significantly increased the amount of data I could harvest from a single video without compromising on the quality of that data.

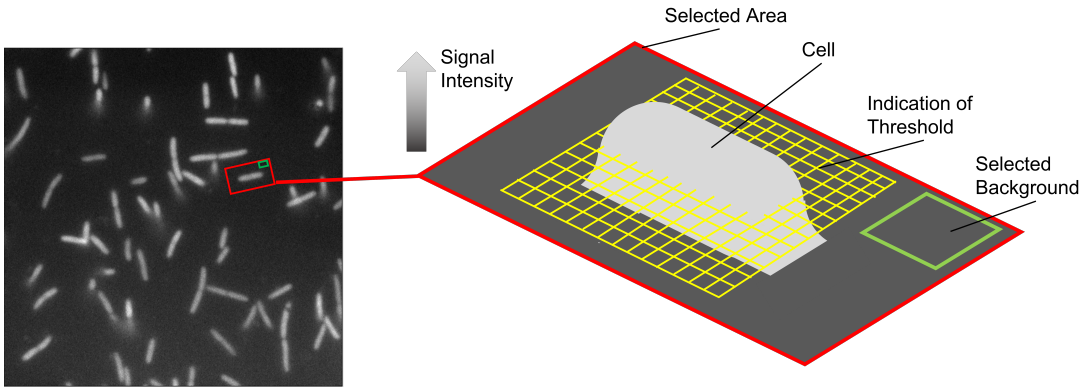


Figure 3.2 *Schematic of the pixel counting method. On the left is a typical FOV containing cells with one good (flat) cell highlighted in the red ROI1. The green ROI2 contains background pixels to be subtracted from ROI1, it should be large enough to get a good average but small enough to reduce sensitivity to passing cells and to keep it away from the cell of interest. The yellow grid represents the threshold above which cells are counted.*

3.2.1 Counting Pixels Above a Threshold

Initially I was working with a program written in LabVIEW by Pilizota, henceforth referred to as the 'pixel counting' method (figure 3.2). This method would take a region of interest (ROI1, red in figure) around a cell with a smaller region of interest (ROI2, green in figure) within ROI1 of only background pixels. The mean of ROI2 is subtracted from ROI1, which is then normalised by the the most intense pixel within it. The normalised pixels which are above a threshold of 0.3 are counted, providing an area count for that cell in that frame. This process is repeated for each frame of each cell in a video. Since the background subtraction occurs each frame this method is fairly robust against additive noise in the whole image which was beneficial as I found there were visible fluctuations in total grey level between frames. One can in principle use the size of a pixel (80 nm^2) to approximate the true size of the cells, however for purposes of reporting changes in area this was unnecessary.

Unfortunately the pixel counting method is not robust to changes in either ROI1 or ROI2 that are not simply noise-based. If a detached cell floats through either ROI it will be incorrectly reported as an area change as the number of pixels above threshold is artificially increased. This can be accounted for post-analysis, but usually the method to do so is to discard the associated cell in ones data store. Crowded cells too are very hard to study with this method, if another cell enters

ROI1 the resulting trace must be discarded as it cannot be said whether any effect comes from either cell. This can be partially accounted for by taking only a portion of a cell in ROI1, but this is not ideal and often leads to confusing or spurious traces. Additionally lowering the number of available pixels by cutting cells in this way decreases the signal to noise ratio, obfuscating any area effects due to phage. As such, if a particular slide has many cells that have clustered, only a few will be analysable per FOV.

It was required then, to generate a method for sizing cells that was robust to crowding and detached cells, and other potential changes in the FOV that were not due to the intensity of the measured cell. An additional benefit was also expected in that the pixel counting method was often very noisy, and developing a new technique may offer ways to reduce this. Given my existing knowledge and its strong capability for numerical analysis I felt it would be easier and more valuable in the long-term to use MatLab rather than LabVIEW to develop further methods. Since I am more proficient in the former any subsequent additions or changes would be much easier to implement, and I could take advantage of the large image analysis libraries that have been developed.

3.2.2 Developing a Method to Detect Edges of Single Cells

In Matlab I wrote a graphical user interface (GUI) to replace the LabVIEW sizing program in my analysis pipeline. I decided to keep using the LabVIEW program which selected ROIs as re-writing this in MatLab, while quite possible, would not represent an increase in usability or adaptability that would reflect the time spent on it. At first I replicated the pixel counting method exactly, so that I had a sensible baseline to compare with. I then tried to simply reduce the noise by normalising ROI1 by the average intensity of the top 10 and 100 most intense pixels rather than just the top pixel. I had anticipated that this would be far more robust to random noise across the ROI, however in practice it reflected a very minor smoothing of the trace and I quickly moved on. I decided to use an edge-finding algorithm to build a contour around a cell and count the number of pixels found within. This method resulted in a satisfactory noise reduction and was robust to detached cells floating through ROI1. It was also able to analyse crowded ROIs if the cell of interest was not actually touching any other cell.

I trialled three edge-finding algorithms for analysis of my data, described in Appendix A.1, and settled on the Canny method. I used the Canny method

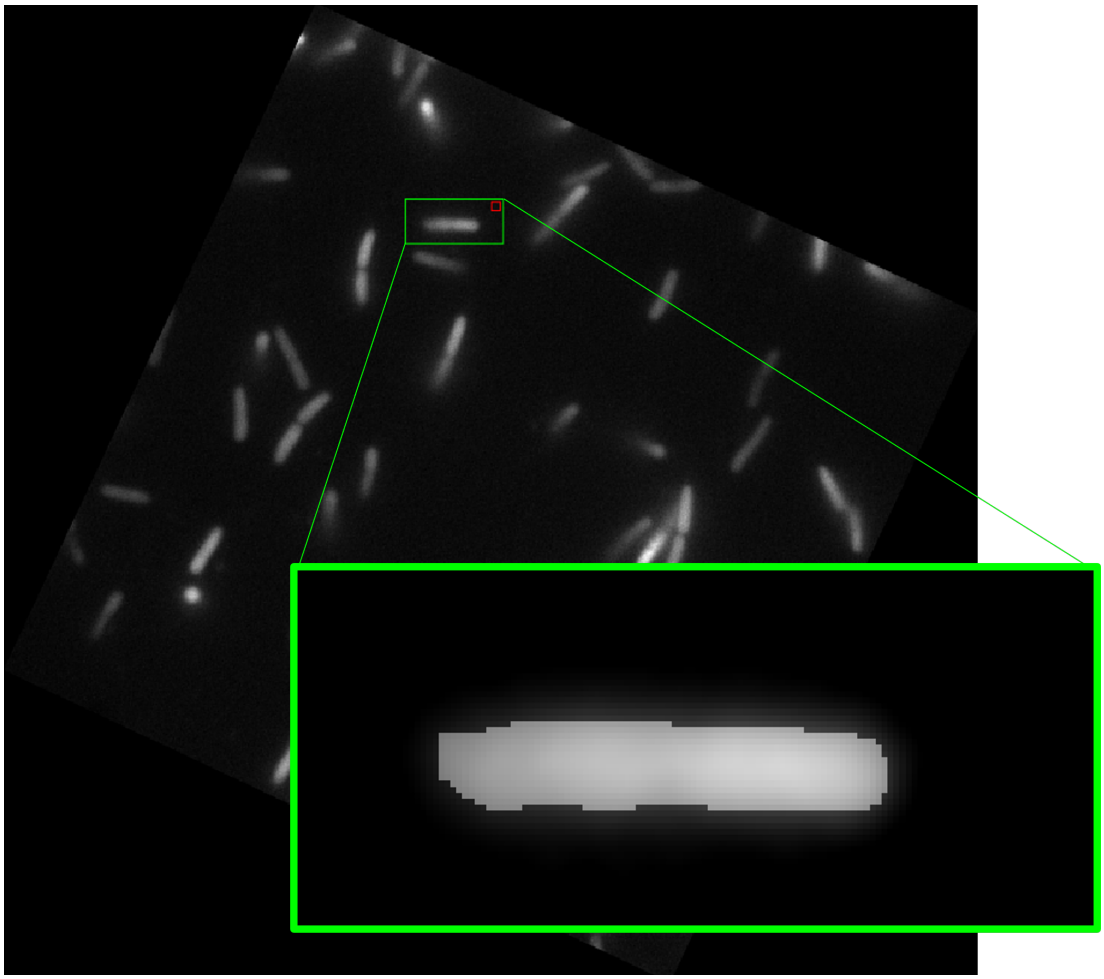


Figure 3.3 *Snapshot of edge-finding analysis of a single cell. The parent image has been interpolated by a factor of 2 to increase the number of pixels along the long axis of the cell, thus decreasing sensitivity to small noise fluctuations along that axis. The algorithm uses two instances of the Canny Edge Detection method, the first generating the threshold for the second which generates the actual binary image. It then passes a set of structuring elements over the binary image before dilating it, filling it, and removing excess noise around the edges using an erosion function. This algorithm slightly reduces noise compared to pixel-counting (approx 20% reduction) and is more robust to other cells intruding on the ROI, however is not capable of addressing closely packed cells.*

to find a contour around a cell, which I then filled to produce cell masks. I then took the masks through a process of erosion and isolation. The isolation ensured that a cell that was intersected by the border of an ROI would not be counted. In this way I did not have to draw clean ROIs around cells to analyse them, and thus increased my data yield from FOVs with crowded cells. While this method was robust to changes in the cell independent of intensity, it could not reliably measure cell size in the event of a cell dividing or a two cells lying flat next to each other.

3.2.3 A Machine-Learning Pipeline for the Size Measurement of Growing and Crowded Cells

The Canny method provided a significant gain over simply counting pixels, however maintained a similar problem in that it was unable to analyse cells that were touching. Initially this wasn't a problem, however it became clear that preparing slides on a bench at room temperature was taking the cells out of exponential growth. At the time this was not ideal, and the consequences of this are reported on in Chapter 5. The solution for this was to prepare slides in a hot room at 37°C. This maintained the growth rates but led to another problem in that cells that grow divide, and cells that divide tend to touch. Thus, the percentage of cells in any given FOV that were analysable dropped significantly even for the more sophisticated edge-finding method.

While I could have spent time coding and optimising to solve this problem, time was growing short and we had become aware of machine learning methods that could solve it faster if adopted. In particular, a time-lapse analysis pipeline called DeLTA2 (**De**ep **L**earning for **T**ime **L**apse **A**nalysis) from the Dunlop lab caught our attention [100]. DeLTA2 is a direct improvement on their first neural network pipeline, which was developed to quickly analyse bulk quantities of one-dimensional mother-machine data [77]. DeLTA2 performs the same task but in two dimensions, with the added capability of tracking cell lineages while outputting both cell age and cell area. Since it is capable of cleverly tracking division and is thus inherently robust in the case of closely packed cells it represented a very strong potential improvement over my edge-finding algorithm.

DeLTA2 is built using two instances of U-Net, a fully convolutional neural net built in 2015 [111]. Convolutional neural networks (CNNs) generally consist of an

input layer, a number of hidden layers where operations take place, and an output layer. Traditional neural networks (also known as fully connected networks) will have a weight assigned to each neuron (node in an operation layer) that can be adjusted in accordance with its training. Each neuron in one operational layer is connected to each neuron in the next layer and so forth and performs a linear transformation on its input vector through a weights matrix.

In a CNN the operations are convolutions rather than linear operations and so each neuron in a layer only has a few connections to the subsequent layer. Because of this, they have a greater flexibility in training and have fewer weights per layer - increasing performance in high dimensional inputs such as image data. Generally a CNN will have fully connected layers that serve as an output where the input to the traditional network is the result of any convolutional layers. The existence of fully connected layers however puts a constraint on the dimensionality of the input and are very parameter rich, and so moving to convolutional layers only represents an increase in the flexibility and decrease in computational cost of the network. The architecture of the therefore *fully* convolutional network (FCN) U-Net is seen in figure 3.4.

For application to my data, DeLTA2 needed to be trained on fluorescence images, and so I generated approximately 150 ground-truth (i.e. human-specified) segmentation maps using Cellpose. Cellpose is another neural net trained only to segment images of cells and represented an easy way to quickly generate maps from my images. The images were swept through in ImageJ as Cellpose, while correctly classifying touching cells, would commonly allow segmentation maps to touch, which is only useful if one retains the class information rather than just a binary mask. I simply drew thin black lines between such merged cells to indicate that they had divided. DeLTA2 requires these ground-truth masks alongside the original images and a weight map which it uses to understand cells that are close together, lines are drawn along the middle of the bacteria and where these lines are close borders are drawn based on the cell width. In this way greater emphasis is placed on both the pixels in the center of cells and those that lie directly in-between cells (figure 3.5). In this way DeLTA2 learns about the small gaps that exist between cells that have just divided or are growing in close proximity and accurately measure each cell in the image.

DeLTA2 is trained using a pixelwise-weighted binary cross-entropy loss function, as detailed in the original U-Net literature. It differs from the original training routine in that custom weights are applied to cells, borders, and the background.

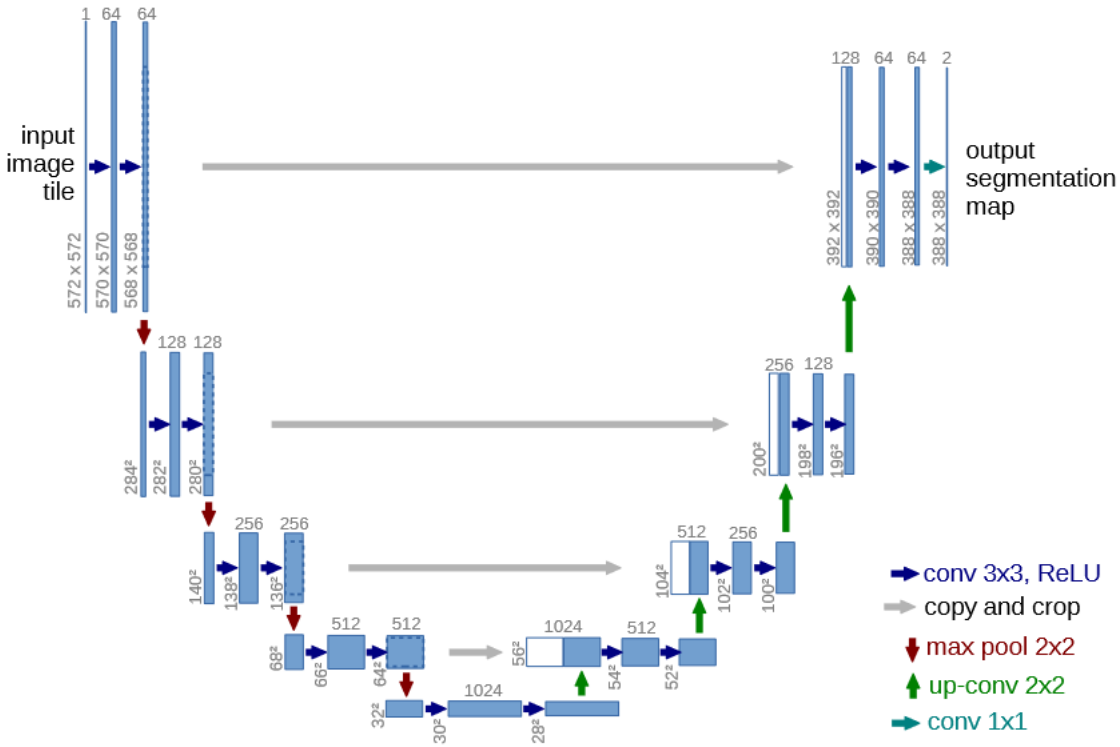


Figure 3.4 *Cartoon showing U-Net architecture. Each blue box is a multi-channel feature map - in the first instance the image to be analysed. Subsequent feature maps (or activation maps) are images that have been produced by the convolutional operations, showing the 'activation' of the image - higher activation occurring in regions where a feature has been identified. The net consists of a contracting path (left) and an expanding path (right). In the contracting path repeated convolutions and subsequent feature pooling downsamples the input images to gain highly efficient classification information. In the expanding path the images are upscaled and compared with images from the equivalent contracting step to retrieve detailed spacial information for the acquired class information. Image taken from [111].*

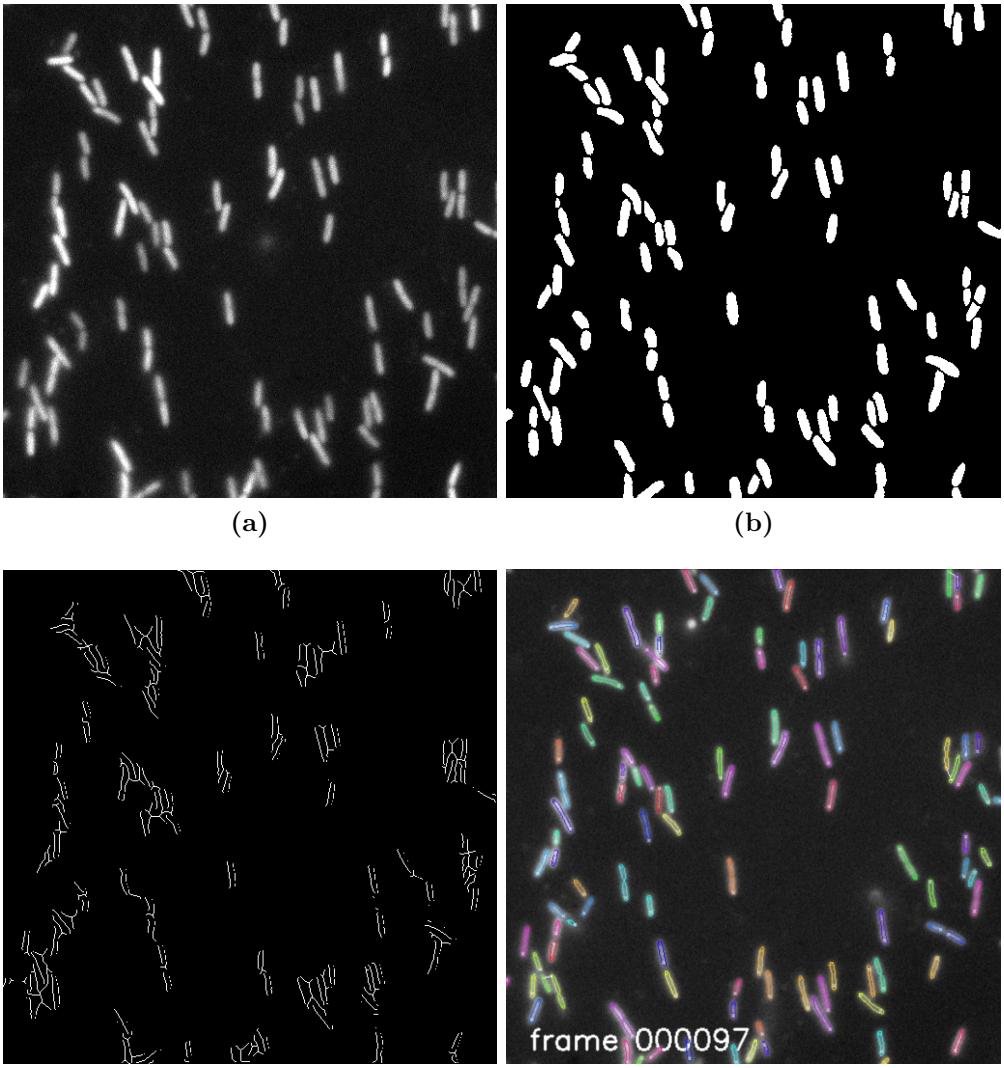
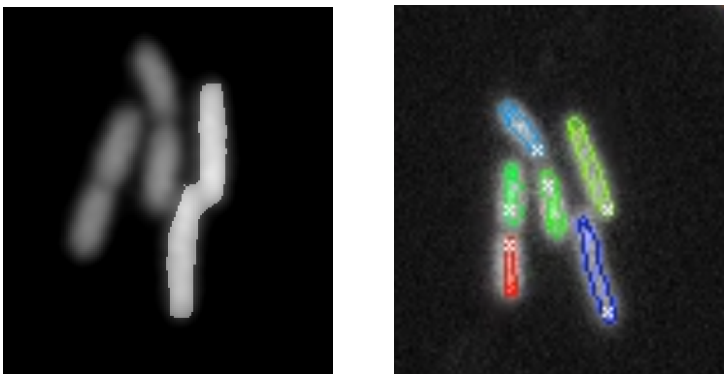


Figure 3.5 *(**a**-**c**) images taken from the training data set for DeLTA2. (**a**) is the raw image, (**b**) is the ground-truth segmentation map produced in Cellpose, and (**c**) is the weightmap which informs DeLTA2 about the small gaps between packed cells. (**d**) the result of running DeLTA2 on a video, the different colours correspond to different instances of cells segmented by the network.*



(a) **(b)**

Figure 3.6 *Comparison of the edge-finding method and DeLTA2 in capability to measure cells in crowded environments. The edge finding method in (**a**) tended to segment more accurately but could not distinguish multiple and had problems with cell-cell joins. DeLTA2 in (**b**) was able to accurately find individual cells but would underestimate their size and shape. Since the feature of interest was a relative change in size this underestimation was not a large problem, however this contributed to high segmentation error to the final data.*

The background is assigned the minimal weight, while cell centers and the borders between cells are given higher weights. Given that the human eye will introduce arbitrary differences in the creation of ground-truth segmentation (if done by hand), lower weights are given to cell-border interfaces to reduce over-fitting.

DeLTA2 is pre-trained on phase-contrast images, and I found that in order for it to understand my data I needed a minimum of four seperate training rounds on fluorescent data. After this there still required some post-processing as there were shortfalls with the output from DeLTA2. In the loading of Flow Cells I rely on the sedimentation of bacteria encouraged by shear flow from syringing over the course of about 15-20 minutes. In spite of high shear syringing many cells are still not completely flat when imaged. DeLTA2 has no problem identifying them, but in this case they are undesirable as they could move around slightly, causing any area trace to be spurious. Since the whole images were generally quite noisy I found methods to exclude traces that involved recognising gradients in the grey values of the bacteria were not very robust and often included or excluded where they should not. With time in mind I simply wrote a script to manually exclude these problematic cells.

A comparison of the efficacy in recognition of individual cells that are lying very close to each other can be seen in figure 3.6. One can see that DeLTA2

has little issue in correctly identifying cells that are touching, whereas my edge-finding algorithm is simply incapable of such. DeLTA2 still had some trouble with dividing cells, and would occasionally erroneously assign mothers and daughters randomly across the FOV. With more time taken to train the lineage tracking portion of DeLTA2 this could have been rectified. However, with time in mind and as the number of such erroneously assigned cells in analysed cells being low it was accounted for manually.

3.2.4 Accounting for High Segmentation Error with Measurement Criteria

From figure 3.6 it is obvious that while the cell identification part of DeLTA2 is strong, it was unable to consistently match the cell size and shape, especially compared to my edge-finding algorithm. Quantifying the segmentation error was done post-measurement, by examining areas of the traces which showed a plateau in growth and were therefore a flat line (shown in figure 5.2 in Chapter 5). By taking the difference between the mean of the plateau and the raw data points it was found that the segmentation error led to a mean-squared error of 8.8%. Here the error is presented as a percentage of the mean of the plateau. This was reduced in python by using the Pandas rolling average function to smooth the trace with a window size of four data points. This way, the error was reduced to just 2.6% without significantly reducing resolution, and any bias introduced by this arbitrary averaging is applied evenly to all traces in all data sets.

This reduction in error was necessary due to a particular criterion for the dips to be counted as real. Explained further in Chapters 4 and 5, there is a relaxed volume of each bacteria which - while it can get smaller through hyperosmotic shock - is about 90% of the cell's initial volume. Therefore any dip that shows a magnitude of more than 10% should not be counted. It was found that adding the small error above to this value did not increase the number of dips found, as the spurious dips tended to have much higher (>20%) magnitudes. A lower bound of 5% was imposed due to the high variation in the segmentation error, 2.6% represents the most common error to see, however it could reach closer to 4.5% in particularly egregious traces, therefore the minimum bound was set so that the window for observing dips was big enough while the lower bound should always be above the error. Since this will undoubtedly artificially decrease the number of dips detected, it would be beneficial for future work to reduce the

un-averaged error as much as possible. Indeed, it should be possible to reduce it to a relatively insignificant value through proper utilisation of the capabilities of the U-Nets upon which DeLTA2 is built.

While DeLTA2 can accurately measure brightfield and fluorescence, it is pre-trained on phase-contrast. Therefore, to run it on fluorescence images as has been done here requires a significant period of training over a large and varied data set. In this thesis, the model was trained to the point of being able to recognise cells and provide a good approximation to their shape, however it would have benefited from being taken further. Strict evaluation of pixel-wise accuracy was not carried out in this case and should be considered by anyone continuing this work.

Alternatively, phase-contrast could have been used to record the images rather than fluorescence, however by the time this was a known problem, the data could not be reproduced. Indeed, additional training could not be carried out due to very tight time constraints at this point in the project.

The segmentation error could also be reduced by employing a marker-controlled watershed segmentation to the images using the DeLTA2 masks as the markers. A watershed transformation essentially treats a monochrome image as a surface where light pixels are 'high' and dark pixels are 'low'. The markers in this case would be the segmentation output of DeLTA2 (or other) for the foreground and the space between for the background. This then informs the watershed function where the maxima and minima should be placed so it can be used to improve the quality of the segmentation. For those interested, there is a rather good explanation on MathWorks, where the process is demonstrated in MatLab. This process was used in the original DeLTA program to produce the ground-truth data training sets for the model, and could be employed to create better masks from the output of DeLTA2.

3.3 Back Focal Plane Interferometry and its Implementation

When I first arrived to the Pilizota lab the microscope that I would be using (named 'Mario') to carry out my experiments was extremely busy. There was another microscope in the room (named 'Bowser') that could be adapted to be

functionally equivalent to Mario. I therefore spent some time performing the necessary adaptations, as this would increase lab throughput and be a useful exercise for understanding one's tools. Missing from Bowser was a system for back focal plane (BFP) interferometry, used in Mario to stabilise the piezo stage and to conduct motor speed measurements on single cells. The optics work on Bowser was completed, however the piezo stage it had did not have the operating range of motion (despite what is given in the technical specification) to be capable of back-focal-plane interferometric stage stabilisation. For this reason, the experiments I performed were carried out on Mario, which is functionally equivalent to what Bowser would have been.

BFP interferometry is a well established method developed to measure forces acting on a particle in the focus of an optical trap. For a bead held in a trap, or tweezer, there will be two contributions of light exiting the trap: the light that interacts with the bead, and the light that does not. The former will scatter in a spherical pattern centered on the bead and interfere with the latter to create an interference pattern. At the BFP of any lens one retrieves the reciprocal of the front focal plane, and as such by imaging the reciprocal of the interference pattern onto a position-sensitive detector (PSD) one may retrieve accurate positional information of the bead. Figure 3.7 shows a schematic of the result of imaging the BFP interference pattern as the bead moves within the trap focus.

The applications Mario and Bowser are designed for do not 'trap' the beads however, instead only pass on the information of the movement of the bead within the focus. This allows the retrieval of positional information without exertion of force on the target bead, and also without any potential harm to biological samples from laser radiation. In order to stabilise the piezo stage against vibrations and small perturbations for long periods of time the PSD is linked to the stage through a feedback loop in LabVIEW. When the bead is moved away from the focus of the laser, the movement is detected and the feedback loop causes the stage to move the bead back to the focus.

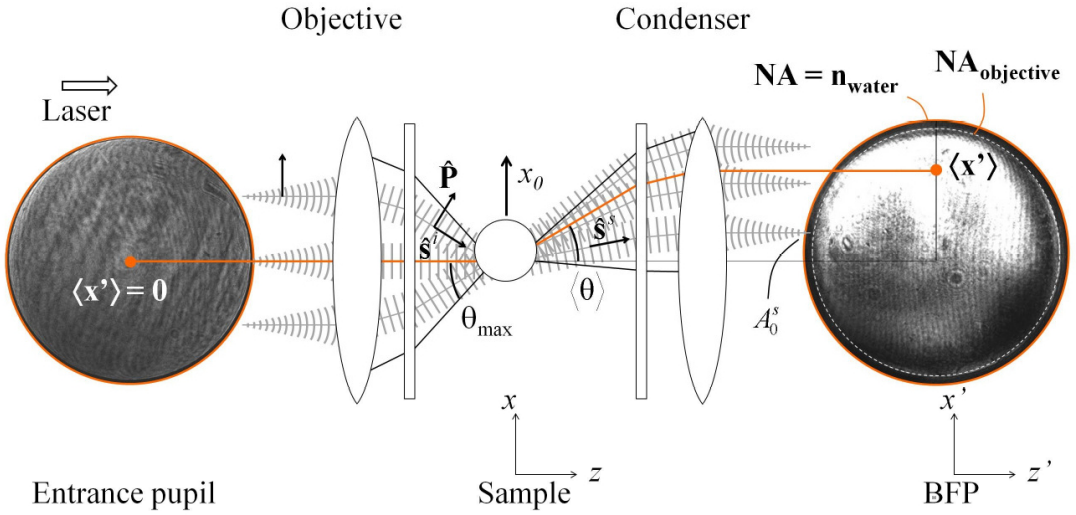


Figure 3.7 *Cartoon showing the result of a bead moving to one side in an optical trap. On the left is an image of the reciprocal of the objective, and the right the reciprocal of the condenser. At $x_0 = 0$ the bead is in the center of the trap and the interference pattern is symmetrical, the two images are identical. If the bead moves then there is also a movement in the distribution of light in the reciprocal of the condenser. Image taken from [33].*

3.3.1 Microscope Adaptation

Theory and Layout

A full schematic of Bowser after implementation of the optical trap can be seen in figure 3.8. The microscope was already capable of imaging at the desired magnification to observe single cells, so in principle all that remained to do was to implement an attenuated optical trap. The trapping laser beam needed to be such that one was able to translate the focus in X and Y independently of Z within the sample plane with relative ease (without explicit knowledge of the microscope alignment). Additionally, the BFP of the condenser had to be imaged directly onto a PSD located above the microscope in a tightly confined space. Thus, as a result of implementing the trap it was necessary to re-align several other systems in the microscope. Most notably this included the bright-field path as the trap necessitated a much shorter working-distance condenser which changed the requirements for Köhler illumination.

I began by planning a layout of optics that would fit around the existing equipment on the optical table. A ray diagram of this layout is seen in figure

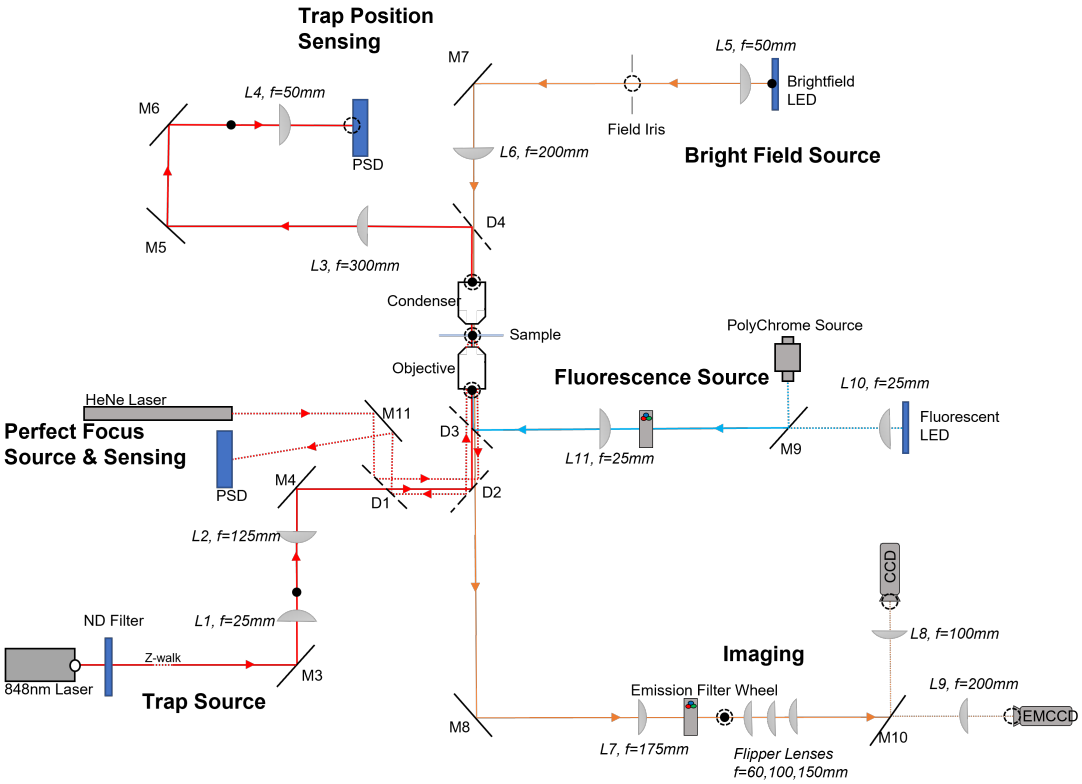


Figure 3.8 *Schematic diagram of the entire microscope. The Trap Source and Position Sensing portions are those I was directly responsible for adding, and I have modified the bright-field for Köhler illumination with the new high NA condenser. The Fluorescence and Imaging portions were in place prior to my arrival. Yellow lines represent the bright-field path, blue lines the fluorescence path, and red lines the 848nm trapping laser used for interferometry. The dashed red line is the He-Ne laser used for a Nikon Perfect Focus system previously stabilising the stage in Z only. The solid black and dashed circles represent conjugate planes, where solid circles are convergences and dashed circles are image planes. The dotted lines on the 848 beam path hide a beam height-correcting Z walk.*

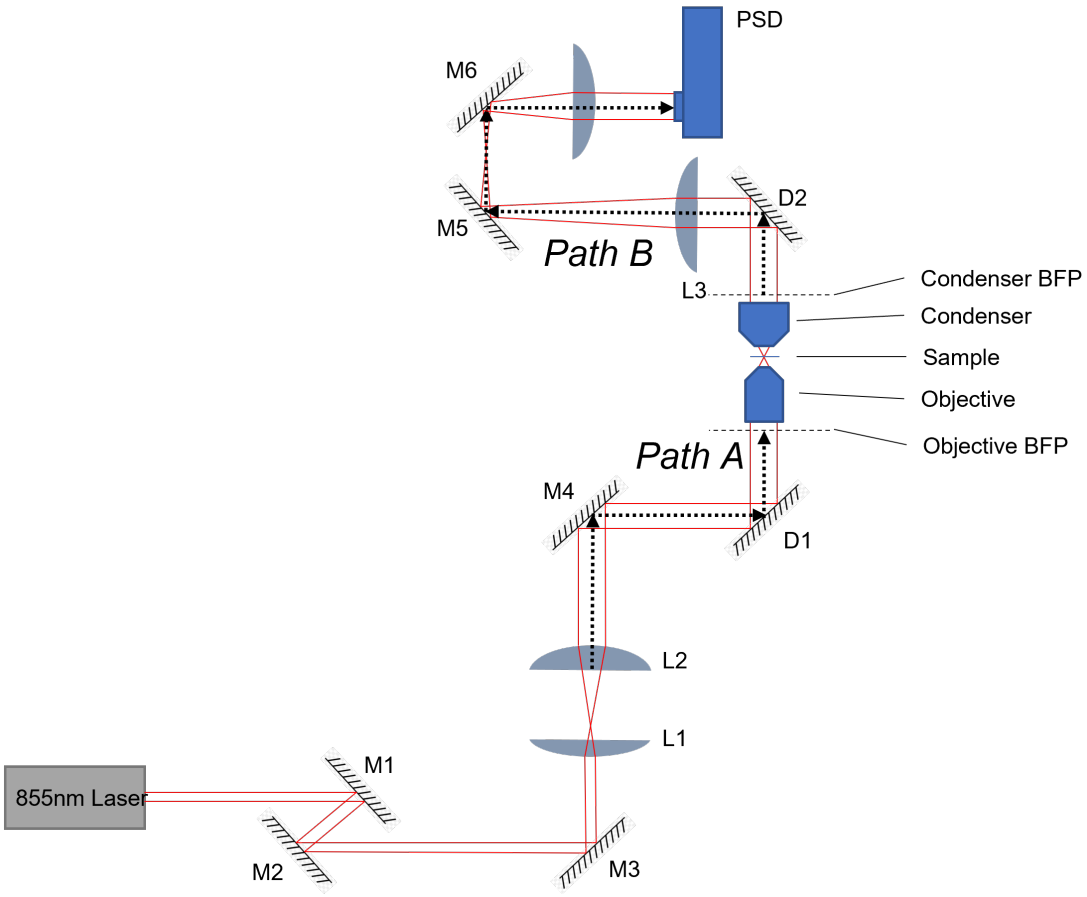
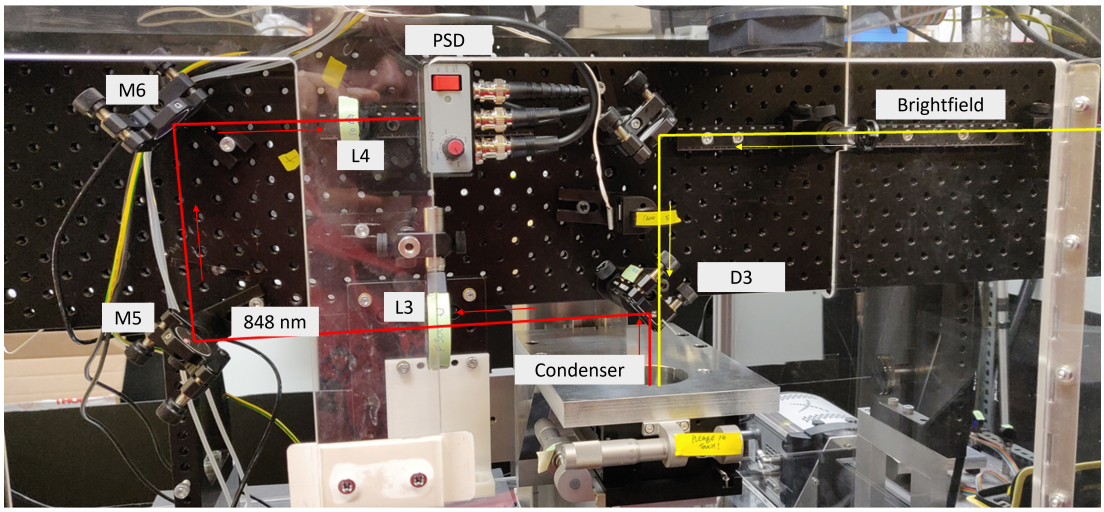


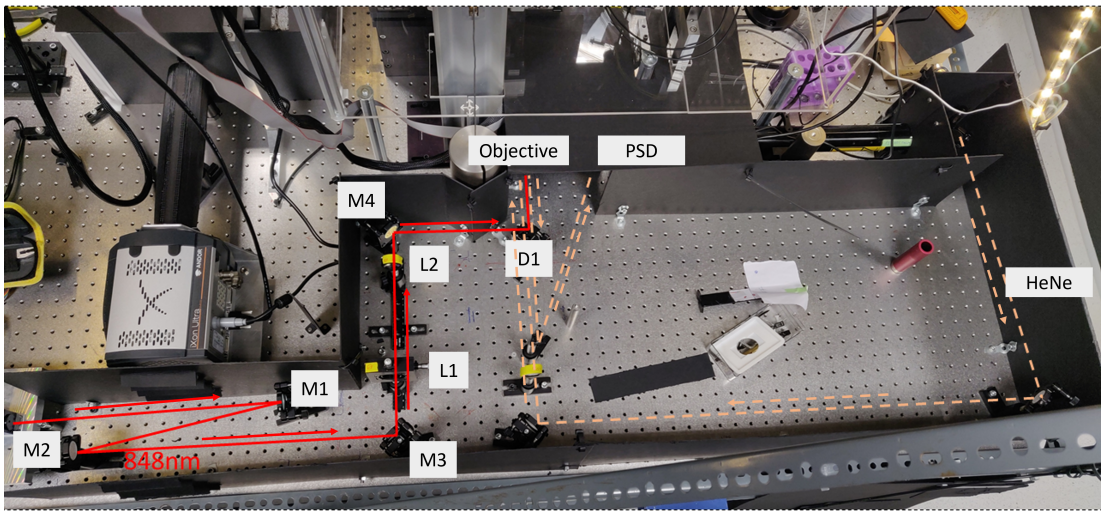
Figure 3.9 *Schematic of plan to implement trap laser. L1 and L2 are separated by the sum of their focal lengths, forming the Keplarian telescope portion. L1 is capable of translating in the plane orthogonal to Path A. D1 and D2 are dichroics allowing the insertion of the laser path into the system without disrupting existing paths.*

3.9. M1 and M2 form a Z-walk that I found was necessary due to a height difference in the laser mount and a mirror under the objective which was difficult to access. The Keplarian telescope portion formed by lenses L1 and L2 fulfills a dual purpose in 1) spreading the laser beam by a factor of $\frac{f_2}{f_1}$ to completely fill the back aperture of the objective; and 2) allowing translation of the laser focus in the sample plane by moving L1 within its plane. In order for this to happen without movements of the lens becoming coupled to changes in the Z position of the laser focus Path A (from L2 to the objective) needs to be such that the telescope's image is in the BFP of the objective.

In general in a two-lens system where light enters lens 1 and leaves lens 2 (figure 3.11), changes in angle of the incident light at a distance x from A are replicated at a distance y from B. To define this distance y I begin with the thin lens



(a)



(b)

Figure 3.10 *Optical paths above* ***(a)*** *and below* ***(b)*** *the objective.* ***(a)*** *light from the 848 nm laser is collected by L3 and L4 such that the BFP is imaged onto the PSD. This allows the stage to be stabilised in XYZ.* ***(b)*** *L1 and L2 form the Keplarian telescope, L1 is on a translation stage that allows movement of the trap in the sample plane. The He-Ne laser (beige, dashed) is reflected off the bottom of the slide and into a PSD allowing stabilisation in* ***Z only****. The 848nm and He-Ne lasers do not operate at the same time.*

equations for both lenses,

$$\frac{1}{f_1} = \frac{1}{s_{o1}} + \frac{1}{s_{i1}} \ , \ \frac{1}{f_2} = \frac{1}{s_{o2}} + \frac{1}{s_{i2}}, \tag{3.1}$$

where f_1 and f_2 are the focal lengths of each lens and s_{o1} and s_{02} and s_{i1} and s_{i2} are the respective object and image distances for each lens.

Here the object of lens 1 (at s_{o1}) is the point of angle change. These will be reproduced as linear movements at a distance s_{i1} from the lens. Since it is desired to replicate the changes in angle at a distance s_{i2} from lens 2, the second lens must use this point that lies s_{i1} from lens 1 as its object. Defining $s_{o2} = d - s_{i1}$ where $d = f_1 + f_2$ is the separation between the lenses, one retrieves for s_{i2}

$$\frac{1}{s_{i2}} = \frac{1}{f_2} - \frac{1}{d - s_{i1}}, \tag{3.2}$$

where

$$s_{i1} = \frac{f_1 s_{o1}}{s_{o1} - f_1}. \tag{3.3}$$

With a little bit of rearranging, Equation (3.2) becomes

$$\frac{1}{s_{i2}} = \frac{d - f_2 - \frac{f_1 s_{o1}}{s_{o1} - f_1}}{f_2 d - \frac{f_2 s_{o1} f_1}{s_{o1} - f_1}}. \tag{3.4}$$

Re-defining s_{i2} and s_{o1} as y and x respectively one finds for y

$$y = \frac{f_2 d - \frac{f_2 x f_1}{x - f_1}}{d - f_2 - \frac{f_1 x}{x - f_1}}, \tag{3.5}$$

Re-inserting $d = f_1 + f_2$ this becomes

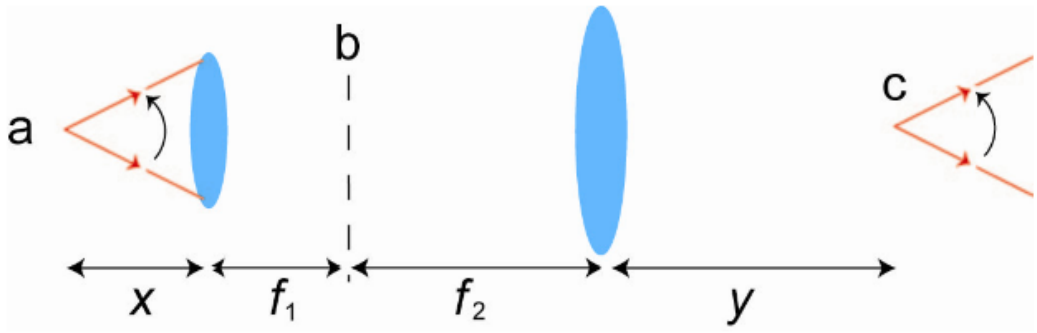
$$y = \frac{f_2^2(x - f_1) - f_1^2 f_1}{f_1(x - f_1) - f_1 x}, \tag{3.6}$$

which becomes

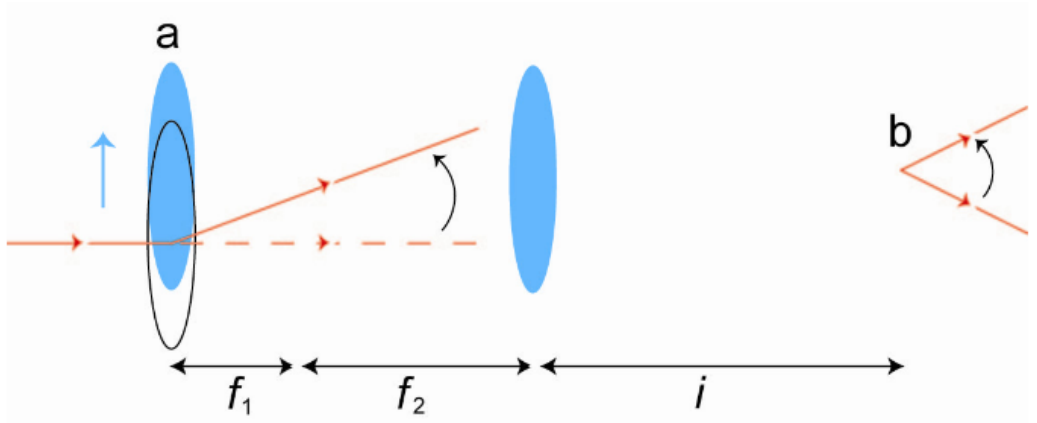
$$y = f_2 - \left(\frac{f_2}{f_1}\right)^2 (x - f_1), \tag{3.7}$$

which holds true if one uses a mirror at distance x from the first lens. However one can simplify this if one chooses to instead translate the first lens instead of using a mirror then $d = s_{o2} = f_1 + f_2$ and the expression for y becomes

$$y = \frac{f_2}{f_1}(f_2 + f_1) = Md, \tag{3.8}$$



(a)



(b)

Figure 3.11 *Cartoon showing the replication of angle change in a beam path at a location further down that path in a two-lens system. (**a**) rotations at **a** are replicated at **c** in this system of lenses. (**b**) the same replication can be obtained by moving lens 1 rather than rotating the incoming beam. (**b**) is experimentally more favourable as a lens in a translation stage is simple to operate and requires fewer components and bench space than if a mirror was used.*

where M is the magnification of the telescope which was previously set for beam expansion. In this case Path A was found to be 750 mm. Path B is similar but simpler, as the L3 and L4 telescope only needs to be such that the condenser BFP is f_3 from L3 and the PSD is f_4 from L4 where L3 to L4 is $f_3 + f_4$.

Re-aligning the Illumination Path

In order for the BFP. interferometry system to function both the objective and condenser need to be low working-distance and high NA. Implementing the laser path thus meant exchanging the existing low NA, high working-distance condenser lens for a new one. While bright-field imaging was not a priority for me, I was not the sole-user of the microscope and so re-aligned it according to the principle of Köhler Illumination[63]. The purpose of such is to avoid an image of the light source (in this case a white LED) of one's microscope in the sample plane by ensuring that this light is collimated. Thus, the image of the LED is de-focused in the sample plane and its conjugates and one achieves uniform lighting across the sample. This is achieved with the presence of an additional collector lens placed before the condenser such that the BFP of the collector is conjugate with the sample plane. A field stop aperture is placed at this point in the collector's BFP such that when the aperture is closed it is visible and in focus when the condenser lens is in the correct position.

In practice, space was somewhat limited and the presence of pre-existing microscope parts meant the collector lens could not be placed in its ideal position. Therefore I achieved Köhler illumination with a Keplarian telescope as the collector with the field stop aperture placed between the lenses at a distance of f from each. Since the aperture plane is conjugate with the BFP of the telescope itself, it fulfills the condition for Köhler illumination. The advantage of using a telescope is that one is able to magnify the LED path to overfill the condenser and ensure proper uniform illumination, despite the added space cost of an additional lens. The distance that the second lens should be from the condenser is then given by (6).

Nikon Perfect Focus

Another system that was in place when I arrived was a Perfect Focus system using a helium-neon laser (HeNe) to stabilise the stage in Z. I could not use this

for my fluorescence experiments as it had no X-Y stabilisation, however I had to move some mirrors and thus spent some time re-aligning this system too. The principle behind the Nikon Perfect Focus system is that a laser incident into the objective reflects off the base of the microscope slide back into the objective and then sent to a PSD. By implementing a feedback loop based on the voltages reported by the PSD one may keep the stage stable in Z. In practice we use an attenuated HeNe laser for this as they are relatively inexpensive and the red light is unlikely to cause photo-damage with any evanescent field created as it is reflected off the bottom of the sample. A half-mirror takes the light exiting the objective to the PSD. The Perfect Focus system cannot be used alongside the trap (it is also unnecessary to do so), and so a mirror on a pop-up mount was used to differentiate the two paths.

3.4 Discussion and Concluding Remarks

The majority of time spent during my program was in the development of the above protocols and analysis methods. There is some discussion already on the measurement protocol in Section 3.1.1, to produce videos of single cells during phage infection, but in general there are several areas for improvement. Possibly the most obvious and easy to improve is the low-throughput nature of the cell imaging protocol. Each experiment produces only a single video, and since each video contains a number of cells (sometimes large, sometimes small) that cannot be analysed. Therefore, a better protocol would use only a Z-stabilised system (e.g. Bowser's Perfect Focus) to take many different FOVs at once, as Mario was inherently limited to a single FOV per experiment due to only having the option of simultaneous XYZ stabilisation. This was the initial motivation for adapting Bowser, alongside the additional proton-motive force measurements that a trap that is not being used for stabilisation could provide. However, this adaptation work was not completed due to at-the-time un-resolvable translation stage issues, and so the protocol had to remain low-throughput. As it stands, I obtained images from 5 different FOVs for the Heat Shock set and 7 FOVs from the No Shock set. This corresponded to 26.7±1.7 analysable cells per FOV for the No Shock set and 21.2±1.3 cells per FOV for the Heat Shock set. One area that this protocol can improve in is to increase this number, largely by improving cell adhesion as only cells that were lying perfectly flat could be analysed. The source of the difference between data sets is unknown, however could be due to some

un-determined temperature effect interfering with adhesion.

Analytically, the methods that were used were changed and adapted as and when it was necessary to do so. Previous work in the lab had been done using the Pixel Counting method (Section 3.2.1) which had been sufficient for those experiments. The key difference between the work presented in this thesis and the work analysed via pixel counting is timing. Osmotic shock can be induced in a matter of seconds, and the recovery measured over the next minutes - the cell does not divide or even grow during this time as its growth is interrupted by the shock. Introducing phage to a growing population of cells and waiting for them to be infected allows the cells to grow and subsequently either divide or elongate into each other.

An ideal scenario would have been planning for the worst and beginning with the most effective method - using DeLTA2 (Section 3.2.3). However, things are rarely so obvious in the thick of it and so various methods were developed and trialed before the machine-learning method was found. The high segmentation errors presented in this thesis are very much the result of the subsequent lack of time for training the model, and indeed exploring other models or functions within the same model.

Given that the analysis pipeline using DeLTA2 or other suitable machine-learning system allows for fast analysis of the data, a clear route to improving this protocol for further study is to increase the number of data points from each experiment. Using a perfect focus Z-stabilisation system as in Bowser would be ideal, such that many videos can be taken from one slide. Then, improvements to the segmentation accuracy of the analysis would allow greater insight into the true statistics of the volume dips.

In conclusion, the sum total of the methods development is an experimental protocol and a computational pipeline that was able to, with a mean error of approximately 2.6%, measure the area of single *E. coli* cells adsorbed to a microscope slide during phage infection. Much work was done to overcome several challenges as they appeared, and there is much more work that can be done to optimize this process. Additionally, work was carried out to adapt a microscope that was unfortunately not able to be completed due to unresolved hardware issues.

Chapter 4

A Mathematical Model of Volume and Solute Loss Driven by Pressure Collapse

As outlined in Section 1.4, a healthy *E. coli* cell maintains turgor pressure in order to grow. The balance of osmotic pressure bringing water in and hydrostatic pressure stretching the membrane keeps the cell inflated according to the Laplace pressure [135]. The consequence of a phage boring a hole into the cell wall will be to disrupt this balance, and the result can be modeled by a relatively simple set of equations. I begin by presenting a simple mathematical model adapted from Buda *et al* [14] which predicts loss of intracellular volume and solutes as a result of a hypo-osmotic shock (environmental osmolarity is reduced). Hypo-osmotic (or down) shock causes the cell to swell as the osmotic gradient acting into the cell is steepened and water flux into the cell dramatically increases. Mechanosensitive channels will open as the membrane is stretched by the swelling allowing solutes and water to escape to equilibrate the osmotic pressure. Since a large volume of water entered the cell some will remain when the osmotic pressure equilibrates and thus there is still an imbalance of pressures driving water out driven by the Laplace pressure. The cell then shrinks below V_0 and must recoup some lost solutes to once again balance the osmotic pressure, slowly drawing water back into the cell and recovering volume to V_0.

In the case of phage infection the process is similar, a hole is opened in the membrane that is assumed to be permeable to water and solutes. The cell should

not initially increase its volume; and should shrink and recover in the same way as above as the cytoplasm is still at a higher osmolarity than the environment. I present an adapted version of the foundational model by Buda *et al* which recovers this result as well as providing insight into the link between hole duration and the magnitude of volume dip.

4.1 The Foundational Model - Volume and Solute Changes During Hypo-osmotic Shock

The pressure/solute model developed by Buda *et al* assumes that an *E. coli* cell is an elastic spherocylindrical shell containing a concentration of solutes such that the internal osmotic pressure,

$$\Pi = -\phi(c_i - c_e)RT, \tag{4.1}$$

is initially at 1 atmosphere (based on experimental values that range from 0.3 atm [25] to 3 atm [16]). Here ϕ is the osmotic coefficient which is set to 1 as the intracellular solution is near ideal. The terms c_i, c_e, R and T are intracellular and extracellular solute concentration, ideal gas constant, and temperature respectively. As described in Section 1.4, the osmotic pressure acts to draw water into the cell which strains the elastic membrane. From the Laplace pressure, acting across a curved surface, one retrieves for the cell-wall tension

$$\sigma = P \cdot r, \tag{4.2}$$

where P is the pressure acting on the cell wall and r is the radius of the cylinder. The model only takes into account the elasticity of the cell wall as it is assumed to be much greater compared to that of the membrane.

Water will flow across the membrane according to the difference in the osmotic and hydrostatic pressures. The flux density across the membrane j [mol.m^{-2}.s^{-1}] is thus given by

$$j = \frac{K(-\Pi - P)}{A_c}, \tag{4.3}$$

where K [mol.Pa^{-1}.s^{-1}] is an 'effective conductivity' coefficient of proportionality and A_c is the superficial area of the cell available for water transit. A_c is kept constant as it is assumed that water flows mainly through porins and that the

total number of porins is constant. For simplicity it is assumed that there is no incompressible volume in the cell, and so volume becomes analogous to water content. The change in volume V over time t can be written as

$$\frac{dV}{dt} = V_m j A_c = V_m K(-\Pi - P), \tag{4.4}$$

where V_m is the molar volume of water. A_c turns out to be largely irrelevant even in the phage consideration as phage infection does not alter the number of existing aquaporins.

In order to fully expand this equation the mechanical pressure must be derived. It has been shown by Deng *et al* that the cell envelope (assumed to be dominated by the cell wall) elasticity in *E. coli* can be expressed as

$$E = E_0 \left(\frac{P}{P_0}\right)^{\gamma}, \tag{4.5}$$

where E_0 and P_0 are the cell wall elasticity and hydrostatic pressure in mechanical equilibrium[25]. The exponent γ was found to be 1.22, however in the model is set to 1 for simplicity. Since the elasticity of a cell wall of thickness l can be linked to the stress by the differential form of Young's Modulus

$$E = \frac{Stress}{Strain} = \frac{d\sigma/l}{dr/r}, \tag{4.6}$$

it follows that

$$\frac{E_0 l}{P_0} \cdot \frac{dr}{r^2} = \frac{dP}{P}. \tag{4.7}$$

Rearranging and integrating equation 4.7 with respect to the change from P_0 and r_0 to P and r gives the expression

$$\frac{\sigma}{\sigma_0} = e^{\frac{E_0 l}{P_0} \cdot \frac{r - r_0}{r_0 r}}. \tag{4.8}$$

Although the pressure here assumes an infinite cylinder, the volume in the foundational model is approximated to a spherocylinder which has the expression

$$V = \frac{10}{3}\pi r^3, \tag{4.9}$$

which when rearranged for r gives

$$r = \left(\frac{V}{\frac{10}{3}\pi}\right)^{\frac{1}{3}}. \tag{4.10}$$

Combining this with equation 4.8 and noting that similar expressions exist for V_0 and r_0 one retrieves

$$\frac{\sigma}{\sigma_0} = e^{\sqrt[3]{\frac{10}{3}\pi}\frac{E_0 l}{P_0}\cdot\frac{V^{\frac{1}{3}}-V_0^{\frac{1}{3}}}{V_0^{\frac{1}{3}}V^{\frac{1}{3}}}}, \tag{4.11}$$

which expands into

$$\frac{P\cdot V^{\frac{1}{3}}}{P_0\cdot V_0^{\frac{1}{3}}} = e^{\sqrt[3]{\frac{10}{3}\pi}\frac{E_0 l}{P_0}\cdot\frac{V^{\frac{1}{3}}-V_0^{\frac{1}{3}}}{V_0^{\frac{1}{3}}V^{\frac{1}{3}}}}. \tag{4.12}$$

In initial conditions the system is in mechanical equilibrium and thus the osmotic pressure and the hydrostatic pressure are equal. This gives for P_0

$$P_0 = \phi\cdot\Delta c_0 RT, \tag{4.13}$$

where $\Delta c_0 = c_{i,0} - c_e$ is the difference in intracellular and extracellular solute concentration that leads to an equilibrium pressure P_0 of 1 atm (approximately equal to the real turgor pressure of *E. coli*). Following from Equations (4.12) and (4.13) we obtain for P the expression

$$P = e^{\sqrt[3]{\frac{10}{3}\pi}\frac{E_0 l}{\Delta c_0 RT}\cdot\frac{V^{\frac{1}{3}}-V_0^{\frac{1}{3}}}{V_0^{\frac{1}{3}}V^{\frac{1}{3}}}}\cdot\Delta c_0 RT\frac{V_0^{\frac{1}{3}}}{V^{\frac{1}{3}}}. \tag{4.14}$$

With now an expression for the hydrostatic pressure one may write equation 4.4 explicitly as

$$\frac{dV}{dt} = V_m K RT\left[\left(\frac{n_i}{V} - c_e\right) - e^{\sqrt[3]{\frac{10}{3}\pi}\frac{E_0 l}{\Delta c_0 RT}\cdot\frac{V^{\frac{1}{3}}-V_0^{\frac{1}{3}}}{V_0^{\frac{1}{3}}V^{\frac{1}{3}}}}\cdot\Delta c_0 RT\frac{V_0^{\frac{1}{3}}}{V^{\frac{1}{3}}}\right]. \tag{4.15}$$

Equation 4.15 explicitly models volume change as a function of the difference in solute concentration across the membrane, a process mediated by the opening of MSCs as stated above. The efflux of solutes through MSCs is the sum of two components, one proportional to mechanical pressure P and one diffusive. One can therefore write directly

$$\frac{dn_i}{dt} = AV_mK \cdot (-P) \cdot \frac{n_i}{V} - Aj_sS, \tag{4.16}$$

where n_i is the number of intracellular solute molecules, and j_s is the diffusion flux of solute molecules per surface area S. S and j_s are defined as

$$j_s = D_s \cdot \frac{c_i - c_e}{l_M}, \tag{4.17}$$

where D_s is the diffusion coefficient of solutes across a membrane of thickness l_M, and

$$S = N_{MSC} \cdot \alpha_{MSC}, \tag{4.18}$$

where N_{MSC} is the total number of MSCs, each with an area of α_{MSC}. Combining equations 4.16, 4.17, and 4.18 one retrieves

$$\begin{aligned}\frac{dn_i}{dt} = &- AV_mK \cdot \frac{n_i}{V} \cdot e^{\sqrt[3]{\frac{10}{3}\pi}\frac{E_0 l}{\Delta c_0 RT} \cdot \frac{V^{\frac{1}{3}} - V_0^{\frac{1}{3}}}{V_0^{\frac{1}{3}} V^{\frac{1}{3}}}} \cdot \Delta c_0 RT \frac{V_0^{\frac{1}{3}}}{V^{\frac{1}{3}}} \\ &- AD_sN_{MSC}\alpha_{MSC} \cdot \frac{\frac{n_i}{V} - c_e}{l_M},\end{aligned} \tag{4.19}$$

where A is a constant denoting the 'openness' of the membrane to solutes and is positive while the MSCs are open and zero when they are closed.

Additional to this is the matter of the active transport of solutes into the cell, a process that is generally always occurring to some degree in growing bacteria. In the foundational model this active transport is assumed to 'switch on' once the volume dips below a threshold value V_{AT} and has a rate proportional to the difference between V and V_{AT}. Thus the flux of solutes from active transport is written

$$\frac{dn_{AT}}{dt} = K_{AT} \cdot \left(1 - \frac{V}{V_{AT}V_0}\right), \tag{4.20}$$

where K_{AT} is a constant of proportionality that is computationally fitted. V_{AT} is assumed to be $0.99V_0$. The summation of equations 4.19 and 4.20 was used to generate the fit seen in figure 4.1.

4.2 Adaptations of the Model for Phage Infection

The situation of phage infection differs from that of hypo-osmotic shock in that the external solute concentration is not changed. In principle this does not require

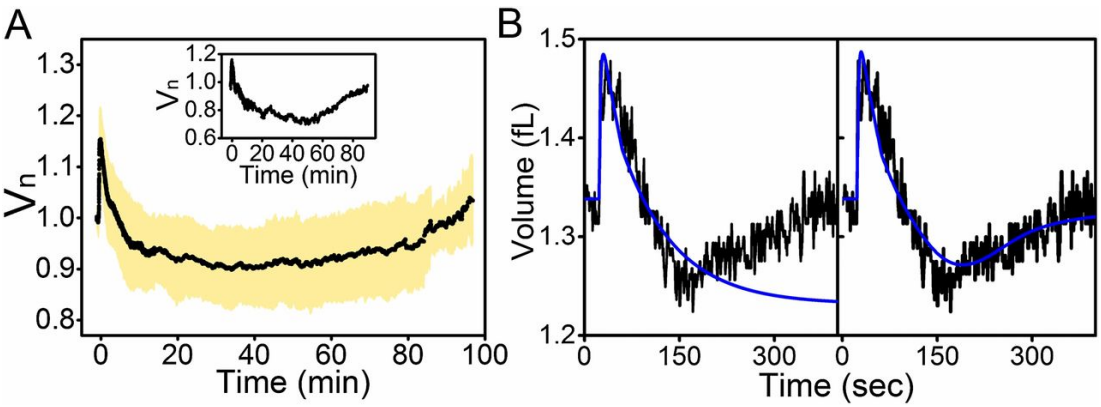


Figure 4.1 *The active response of cells to hypo-osmotic shock. (**A**) The black line is the average volume vs time for 36 wild-type strains exposed to osmotic downshock. (**B**, left) A representative volume trace from a single cell (black), with a fit (blue) produced by Equation 4.19. (Right) The fit performed with the active component in Equation 4.20 included. Image taken from [14].*

any particular change to the equations themselves. One can simply open a hole by explicitly setting A to be positive in equation 4.19 and reducing the area to match that of a single phage hole rather than a host of MSCs. Rather than the diffusive term in equation 4.19,

$$\frac{dn_i}{dt}_{diff} = -AD_sN_{MSC}\alpha_{MSC} \cdot \frac{\frac{n_i}{V} - c_0}{l_M}, \tag{4.21}$$

one would write

$$\frac{dn_i}{dt}_{diff} = -AS_pD_p \cdot \frac{\frac{n_i}{V} - c_0}{l_M}, \tag{4.22}$$

where S_p is the surface area of the phage hole and D_p is the diffusion coefficient of solutes through that hole. In the first instance of this model D_p is assumed to be equal to D_s, and it is assumed that the number of phage infecting the cell is 1. The original model collapses several terms into a fitted parameter $\alpha = \frac{N_{MSC}\alpha_{MSC}D_s}{\Delta l}$ which is computationally fitted as 0.74632. From this, D_s (and therefore D_p) can be estimated to be 3.2×10^{-4} nm^2.s^{-1}.

The original model had a simple rate process where import is governed by a rate constant. The active pumping is dependant on V and only turns on when $V < V_{AT}$ where V_{AT} is the associated threshold for active transport and is set at $0.99V_0$. However, in a real cell the transport of solute is mediated by the balance of leakage and import, a detail that is omitted by the foundation model and as such the active import is one of its more made-up parts. As such, a more complete version of equation 4.19 used in the phage-adapted model is

$$\begin{aligned}\frac{dn_i}{dt} =& - AV_mK \cdot \frac{n_i}{V} \cdot e^{\sqrt[3]{\frac{10}{3}\pi}\frac{E_0 l}{\Delta c_0 RT}\cdot\frac{V^{\frac{1}{3}}-V_0^{\frac{1}{3}}}{V_0^{\frac{1}{3}}V^{\frac{1}{3}}}} \cdot \Delta c_0 RT \frac{V_0^{\frac{1}{3}}}{V^{\frac{1}{3}}} \\ & - A\frac{S}{V}D_p \cdot \frac{\frac{n_i}{V} - c_0}{l_M} + Import - Leakage.\end{aligned} \tag{4.23}$$

E. coli has a few ways of importing nutrients, the most important of which are the ATP driven pumps (the transcripts for which make up 5% of the genome)[94]. From a different model presented by Terradot *et al* [129], wherein the proton-motive force is linked to the regulation of cytoplasmic pH, one retrieves for solute transport through membrane channels

$$Import = j_+(1 - e^{-\frac{F}{\rho R}\Delta G_{n_i}}), \tag{4.24}$$

where ΔG_{n_i} is the electrochemical potential of a solute n_i:

$$\Delta G_{n_i} = z_{n_i}\Delta\Psi + \frac{RT}{F} ln \frac{c_i}{c_e}. \tag{4.25}$$

The term z_{n_i} is the solutes valency (assumed ± 1), $\Delta\Psi$ is the membrane potential, F is the Faraday constant, and j^+ is a function of the transport kinetics. In this model I consider only non-charged solutes that travel through ATP driven pumps. I therefore assume that ΔG_{n_i} is very large as exponentially growing cells have a wealth of intracellular ATP and it should not be immediately depleted on phage infection, so that $e^{-\frac{F}{\rho R}\Delta G_{n_i}} << 1$ and $Import \to j_+$. In this way it is assumed that when a solute contacts a channel it will pass through in a single step with a probability of 1. Importantly, if the growth of the cell is interrupted this assumption may no longer be true and this model will need to be adapted to account for ΔG_{ni}. The function j_+ is dependent on the kinetics of import which is assumed to be a single step process. Assuming that *Import* must balance *Leakage*, one may retrieve a function for j_+ when $\frac{dn_i}{dt} = 0$.

Assuming that the leakage is driven by the osmotic gradient, one can write it similarly to the previously defined diffusive component of $\frac{dn_i}{dt}$. However, importantly the 'diffusive coefficient' D_{ni} is not a truly diffusive term as the *Leakage* term includes all solute transport out of the cell. It is sufficient for

simplicity however, and the cell's natural efflux of solutes can be written as

$$Leakage = -SD_{ni}(\frac{n_i}{V} - c_e), \tag{4.26}$$

where D_{ni} is the diffusion coefficient (more accurately a 'leakage coefficient') over the surface of the cell S. When a hole is closed, i.e. $A = 0$, the $\frac{dn_i}{dt}$ expression is comprised only of import and leakage and so

$$\frac{dn_i}{dt} = j_+ - SD_{ni}\frac{\frac{n_i}{V} - c_e}{l_M}. \tag{4.27}$$

Which when there is no overall transport of ions (i.e. $\Pi = \Pi_0$) becomes,

$$j_+ = SD_{ni}\frac{\frac{n_i}{V} - c_e}{l_M}. \tag{4.28}$$

The numerator in equation 4.28 is equivalent to $\frac{\Pi}{RT}$, and so the expression may be simplified to

$$j_+ = \frac{SD_{ni}}{l_M RT}\Pi\Big|_{\Pi=\Pi_0} = \frac{SD_{ni}\Pi_0}{l_M RT}. \tag{4.29}$$

Thus the whole form of $\frac{dn_i}{dt}$ when $A = 0$ is

$$\frac{dn_i}{dt} = \frac{SD_{ni}}{l_M RT}(\Pi_0 - \Pi), \tag{4.30}$$

which can be rearranged as

$$\frac{dn_i}{dt} = \frac{SD_{ni}\Pi_0}{l_M RT}\left(1 - \frac{\Pi}{\Pi_0}\right), \tag{4.31}$$

and the active component of ion transport is therefore proportional to the difference in osmotic pressures rather than explicitly dependant on the volume. This is the biggest assumption of the model, and is true only if the active transport of solutes balances a leakage to recover the osmotic pressure back to Π_0. The complete form of equation 4.23, substituting $L = \frac{SD_{ni}}{l_M RT}$, is therefore:

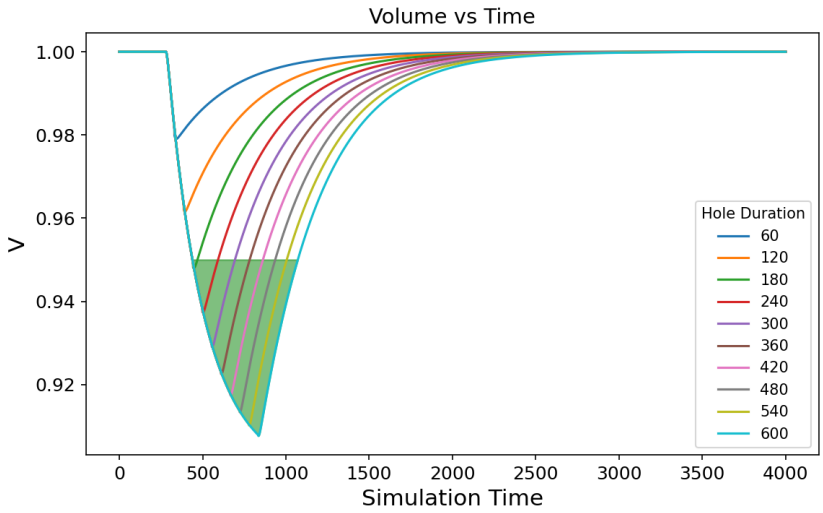
$$\begin{aligned}\frac{dn_i}{dt} =& - AV_mK \cdot \frac{n_i}{V} \cdot e^{\sqrt[3]{\frac{10}{3}\pi}\frac{E_0 l}{\Delta c_0 RT}\cdot\frac{V^{\frac{1}{3}} - V_0^{\frac{1}{3}}}{V_0^{\frac{1}{3}}V^{\frac{1}{3}}}} \cdot \Delta c_0 RT\frac{V_0^{\frac{1}{3}}}{V^{\frac{1}{3}}} \\ & - A\frac{S}{V}D_p \cdot \frac{\frac{n_i}{V} - c_0}{l_M} + L\Pi_0(1 - \frac{\Pi}{\Pi_0}).\end{aligned} \tag{4.32}$$

As shown in figure 4.1 the base model for hypo-osmotic shock is adequate for describing experimentally observed behaviour as the opening of MSCs is regulated by the volume of the cell. However, it has been found experimentally that an *E. coli* cell cannot lose more than 10% of its volume due to structure retention in a de-pressurized cell wall [105]. While the foundational model does not recover this result, the phage adapted model does. However, this was achieved by fitting L such that the volume during the hole open phase does not drop below $0.9V_0$. All parts of L are approximately known apart from D_{ni}, which after fitting was determined to be 2.88×10^{-9} $\mathrm{m}^2.\mathrm{s}^{-1}$, about 5 orders of magnitude smaller than the diffusion coefficient for an open hole in the membrane. As above, D_{ni} is not technically a true diffusion term, rather it accounts for the loss of ions from general cell processes and therefore should be expected to be much lower than D_s. The stop at $0.9V_0$ is not obtained through any definite mathematical rule, rather it was imposed as a condition on the model to fit L. In future iterations of the model it would be better to consider V_0 in terms of P_0 and E_0 where the volume does not decrease too much due to a 'relaxed volume' in the cell, rather than because the import speeds up or slows down depending on the osmotic pressure.

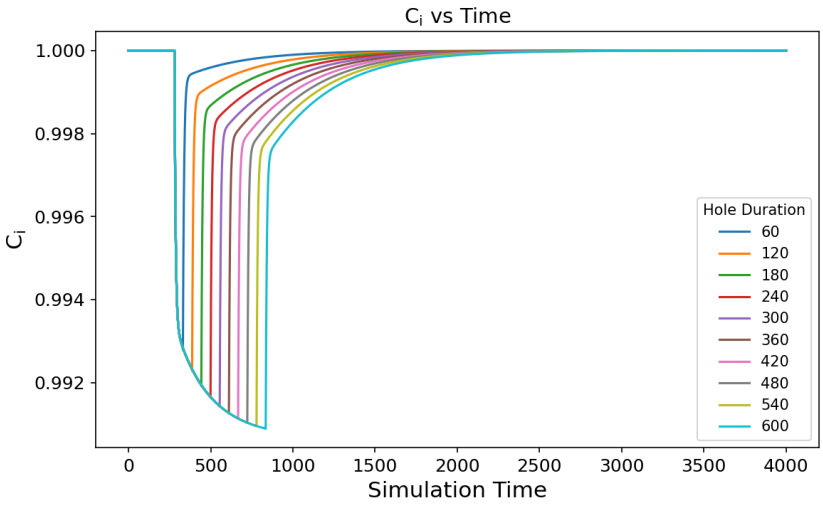
The shaded part of figure 4.2 indicates magnitudes of dips that could be resolved during analysis of the traces. This means that any dip where the minimum volume V_{min} did not reach 0.95 would not have been recognised. Any dip where V_{min} is less than 0.95 is then assigned a duration assuming that the minimum change in normalised volume resolvable is about 0.01. A clearer illustration of this is in figure 4.3, where the relationships between V_{min} and hole duration and dip length are shown.

4.3 Discussion and Concluding Remarks

This model was original developed to describe hypo-osmotic shock, wherein the cell loses volume due to mechanosensitive channels in the membrane opening as it comes under tension from water influx. In the case of a single phage infection, as is assumed, a single hole is made in the membrane for DNA to be trans-located into the cell. The origin of volume loss is therefore similar, however has a few key differences. Namely, the number and distribution of mechanosensitive channels is greater and wider than a single phage hole. MG1655 has been measured to have around 400-1000 MSCs per cell depending on growth phase [10]. These channels are distributed all over the membrane and so simply reducing the value of N may

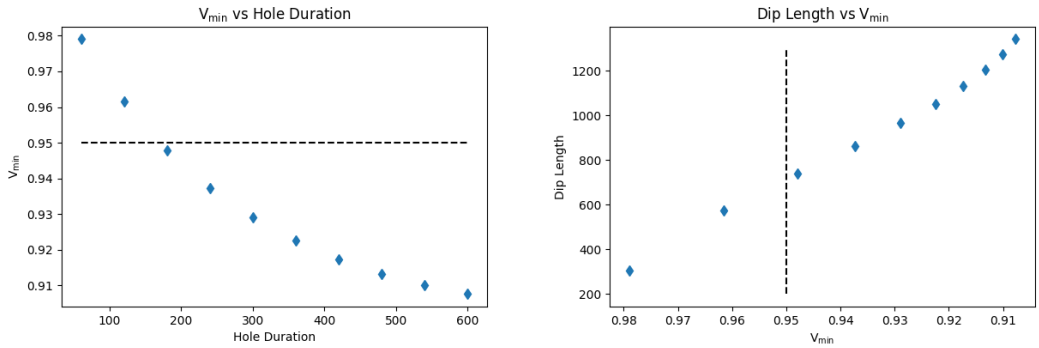


(a)



(b)

Figure 4.2 *The model predicts that as a hole opens in the membrane the (**a**) volume and (**b**) c_i values will drop sharply before recovering. The shaded area marks the bounds of the dip magnitude criteria applied ($> 5\%, < 10\%$, as in Section 3.2.4) in my analysis. The amount of both solute and volume lost is proportional to the duration of the open hole. The concentration of intracellular solute c_i does not drop much at all ($< 1\%$) which mirrors some experimental data which demonstrates a lack of decrease in ionic c_i during T7 infection, however as the adaptations do not take into account specific phage type this is likely a coincidence. Experiments monitoring the size of a bacterium should see a decrease in the size of the cell. The criteria for hole measurement had a cutoff at 5%, i.e. dips of a lower magnitude were not counted. This model predicts that holes may only cause measurable dips if they are open for more than approximately 3 mins, and between 3 and 10 mins the distribution of dip durations should be fairly wide.*



(a) **(b)**

Figure 4.3 *(**a**) the range of V_{min} for the hole durations seen in figure 4.2, the black hashed line showing the cut-off for observable dips. The intersection of the data curve and this line marks the minimum duration then for a hole to be open for the dip to be observable by my analysis. (**b**) the lengths of the dips that have minimum volumes V_{min} for the dips shown in figure 4.2, the black hashed line again showing the minimum volume for observable dips. The intersection of the data curve and this line is thus the hypothetical shortest dip that would be resolvable by my analysis. While the numbers involved are likely inaccurate (I do not see many dips stretching beyond the 8 minute mark, but this suggests anything below 10 should be invisible), the proportionality between dip length and dip magnitude may be equivalent to reality. Given the disparity between measurement and prediction it seems sensible to conclude there may be a faster or functionally different recovery of volume and fitting to experimental data may improve this.*

not be sufficient to account for this difference. It *may* be indeed that the hole created by the phage that is permeable to ions is quite a bit larger than the hole created for DNA transport - which in T7 is very well defined as it is formed of a large protein complex.

As previously mentioned, it is not necessarily believable that the timescales this model reports are realistic. As above, the numerical scaling is done relatively naïvely so one may reasonably expect the absolute numbers to be off. Indeed, it appears as if there should be no measurable dips shorter than 700 s, however in Chapter 5 we find the average duration to be just over 300 s. However, an argument can be made that the *dynamics* of volume loss should be similar if phage infection mirrors osmotic shock in that solutes may exit the cell cause an overall loss of volume. Therefore one might expect the *function* of the volume to be accurate with a high scaling error. If this is true then the model shows that the volume loss is proportional to the duration of the hole, with the limitation of not allowing more than 10% of volume loss being recovered. Given that any analysis method will produce some noise, this model also predicts that for any experiment there is a minimum hole duration for the volume loss to be detectable should the distribution of hole-durations be sufficiently wide. In my experiments there was an approximate 5% signal noise, meaning any change in volume that was less than 5% could not be separated from noise. In Chapter 5 I apply this noise to both a shallow exponential control and to the above traces produced by the model to show that this predicts an uncertainty proportional to experimental conditions.

On the topic of timescales and T7, and there is more discussion Chapter 5, it is important to recall from Section 1.3.2 that T7 DNA is transcribed into the cell. At 30°C this should take about 10 minutes in total at a rate of 47 $bp.s^{-1}$, however rates vary quite dramatically with temperature. At 37°C the rate increases to about 60 $bp.s^{-1}$, and at 20°C the rate decreases down to only 10 $bp.s^{-1}$ [60]. This wide variation means that if volume loss is mediated by the DNA channel only, one would expect a large difference in dip length between different environmental temperatures and almost *no* difference between experiments where the temperature is the same.

This model does not involve any particular mechanisms regarding hole opening and closing. It only assumes that there is a hole formed and that this hole is permeable to solutes. How the hole forms, and indeed how it closes, are unknowns that cannot be known from the results in this thesis. What can be said however

is that, given the functional dependence of the two, the dip duration is unlikely to be smaller than the hole duration, which is realistic given that this would involve incredibly heavy ion pumping from the cell. It is far more reasonable to think that the hole is closed which allows recovery via normal ion pumping, as is assumed in the model. Therefore, this model should provide a reasonable start to the process of numerically understanding the process it describes. Another reasonable increase in utility of this model would be to implement a simulation of the effect of cold shock. Since the biological effect of a longer hole duration is already presented, one could instead look at changing the value of permeability to solutes as this may represent the affect that different temperatures have on the membrane of the cell.

In conclusion, this shows that if the number and size of mechanosensitive channels in a model *E. coli* cell are reduced to account for a single phage-made hole, the dynamics of volume loss and solute transport are conserved. Therefore, one may expect to observe losses in volume from observing single bacterial cells under a microscope when phage are introduced. It also shows that there is a minimum observable volume loss, which means that any observed stochasticity likely arises from a distribution of hole durations.

Chapter 5

Mechanical Observation of Phage Infection of Single *E. coli* Cells

5.1 Phage Infection is Marked by a Drop in Cell Volume

In Chapter 4 it is presented that the volume of a cell infected by a single phage should decrease in a way that is proportional to the open duration of a hole or pore formed by that phage. I present here that from observations of single K12 *E. coli* cells in RDMG that there is a volume loss of 5%-10% total volume that is assumed to correlate with the infection by phage T7. There are two data

	No Shock Set	Cold Shock Set	No Shock Control	Cold Shock Control
N_{total}	170	102	101	64
N_{dips}	67	64	12	3
Dip Probability	39±4%	62 %±7%	17%±4%	8%±4%
Dip Length (min)	3.42±0.31	7.69±1.23	4.60±2.7	5.27±1.31

Table 5.1 *Table of results from measuring 272 cells in total across two data sets differentiated by No Shock and Cold shock where N_{total} is the total number of cells in that data set and N_{dips} is the number of cells which exhibited a drop in volume. Cells in the Cold Shock set were at room temperature (19°-20°C) for 20-30 mins, while cells in the No Shock set were kept at 37°C from incubator to microscope. Dip magnitude is not included here as it was on average approximately equal between all sets. Control sets were not given phage, and all cells in non-control conditions were infected.*

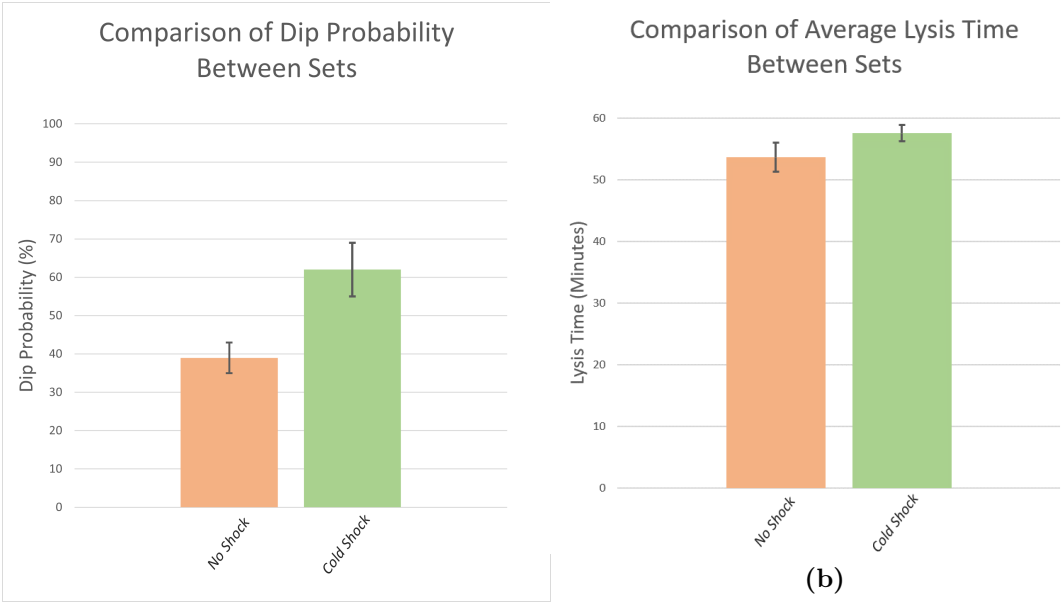
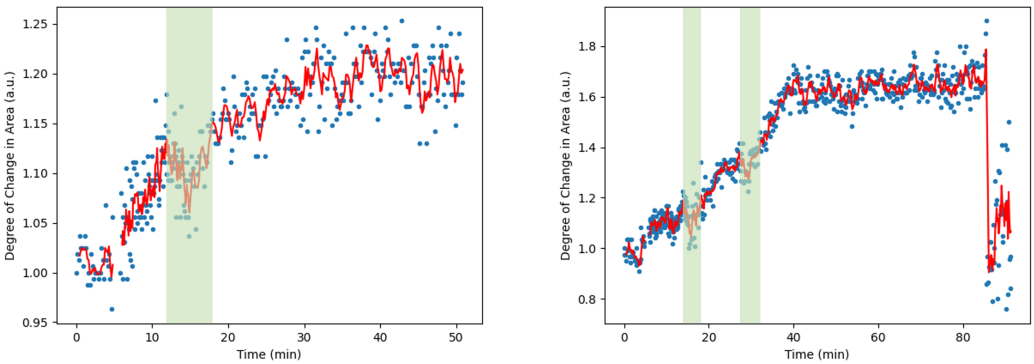


Figure 5.1 *Graphical Representation of Table 5.1, with the addition of a comparison between the lysis times (i.e. infected-cell lifetime) between the two data sets. There is a large difference between the two dip probabilities, and a much lesser difference between lysis times. Overall the Cold Shock set grew slightly slower, so one may expect the cells to take slightly longer to lyse. Due to the closeness of the lysis times between the two sets it is not obvious that this difference should be the sole cause of the disparity in dip probability, but it may be a factor as discussed in Section 5.2.2.*

sets presented as, recalling to Section 3.1.1., there was a difference in experiment preparation temperatures, leading to the two data sets Cold Shock and No Shock. Table 5.1 shows overall results from a study of 286 phage-infected cells, wherein Dip Probability refers to the chance to see a dip of average Dip Length in the associated data set. The Cold Shock set contained 1.6 times the number of volume dips than the No Shock set, and on average those dips were 2.3 times longer. A "bulk" comparison of dip probability and lysis time can be seen in figure 5.1 and representative cell area traces can be seen in figure 5.2.

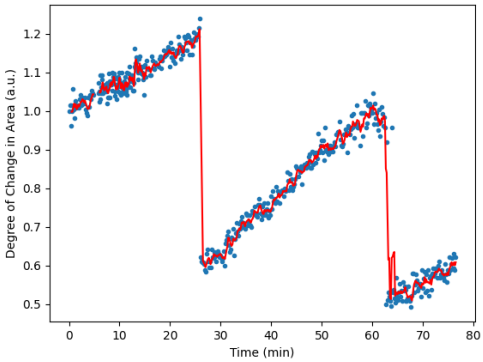
The controls follow the opposite pattern to the phage experiments in that the higher percentage of dips is in the No Shock control set and in the Cold Shock phage set. In both cases the control dips are significantly less frequent, and their presence is likely due to sources of noise in the analysis constructively interfering.

As the Cold Shock cells were resting at room temperature for a short time, their steady-state growth was interrupted. Cells in this set exhibited lower-on-average



(a)

(b)



(c)

Figure 5.2 *Representitive traces from (**a**) set Cold Shock, (**b**) set No Shock, and (**c**) the control set. Raw data points are in blue, while the red line signifies a rolling average taken over four data points. Here the Y-axis shows degree of change in area rather than raw pixel count, this was a choice to improve readability in the graph, it does not scale the trace in Y. Highlighted in (**a**) and (**b**) are the area dips, and one can see the growth of the cells in the phage sets plateauing some time after infection - this is particularly obvious in (**b**). Phage traces that do not show dips are still highly likely to present these growth plateaus. The control trace shows the typical growth - division - growth cycle of an unperturbed bacterium following a doubling time of approximately 35 min, slower than expected for EK03 growing at 37°C. In all traces there is a gap at 5 minutes where imaging stopped to allow syringes to be swapped. Prior to this gap there is no phage present in any condition. Additionally the first 15 minutes of each trace had a higher time resolution, this was to improve the chances of seeing shorter dips as we believed dips may occur more frequently within 10 minutes of phage addition.*

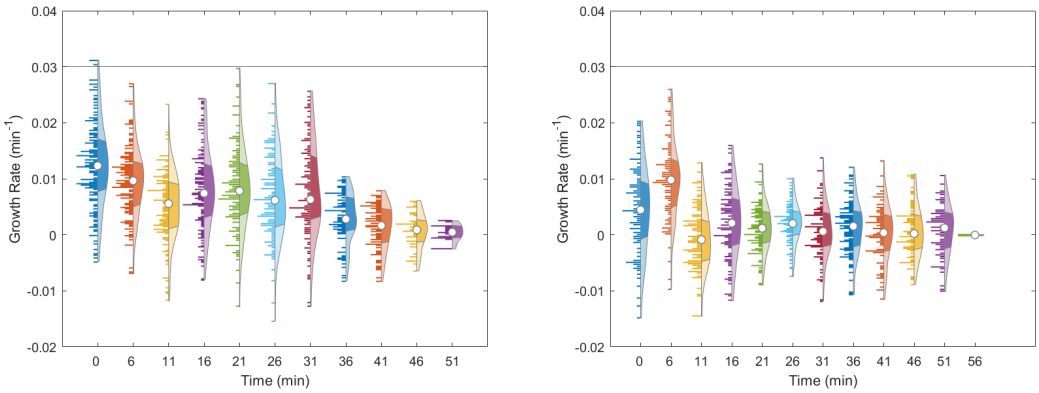


Figure 5.3 *The time-evolution of cell growth rate for the (**a**) No Shock set and (**b**) Cold Shock set. The rates were calculated from a single exponential fit to data points in windows of five minutes along each cell area trace. The horizontal line marks the normal exponential growth rate of EK03 E. coli cells in bulk. The cells that have been left at room temperature report lower-on-average growth rates at the beginning of the infection. In both cases the growth rate tends to zero as the cellular growth plateaus as a result of the infecting phage. Negative growth rates are a combination of the consequence of fitting noisy data and the volume dips which occurred on average at (**a**) minute 13 and (**b**) minute 15.*

growth rates prior to phage addition than their No Shock counterparts (figure 5.3). Additionally, the No Shock cells exhibited a greater number of divisions with 25% compared to only 8% in the Cold Shock set, further highlighting the interrupted growth cycle. In the Cold Shock set approximately 6.7% of dips were seen in dividing cells and in the No Shock set approximately 29.9% of dips were in dividers. Since these are both similar to the raw division numbers there does not appear to be any particular correlation between number of dips and division. Cells that divided in this case were considered only as a mother-to-daughter cell. Ideally the two daughters would be treated as two new cells, however inadequate division tracking from DeLTA2 meant one daughter was almost always lost. Therefore the daughter that remained would be chosen to continue the trace of the mother. In both data sets the division events took place on average before the dips, which one might expect given what we know of the interference with cell growth by T7. Additionally, between the two data sets there was no significant difference between the mean times between dip occurrence and cell lysis, with the Cold Shock set showing approximately 42.9±1.6 minutes and the No Shock set showing approximately 40±2.4 minutes between volume dip and cell death.

Another marker of phage infection observed in both sets is the plateauing of

growth (visible in figure 5.2) some time after infection. This is observed in 95.4% of all cells, and has approximately the same likelihood to be observed between both sets of data. T7 (and phages in the lytic cycle in general) will completely disable the synthesis of host proteins. The lack of difference between plateauing in both sets indicates a stoppage of new biosynthesis regardless of initial growth rates.

It is assumed that the plateaus are due to the phage shutting down cellular processes and halting growth of the cell. However, while all cells that were analysed did lyse, only around 55% of cells showed the plateau before lysing.

5.1.1 A Note on the Criteria of Measurement

In Section 3.2.4 and Chapter 4, figure 4.2, the criteria of measurement of a dip is referenced to be a magnitude loss of between 5% and 10% of it's volume. The lower bound of this was chosen to be high enough over the segmentation error so that the dips would not be measured spuriously. While it is true that *E. coli* can lose up to 60% of its volume without plasmolysis in the case of hyperosmotic shock [104], in this case the cells are not being shocked externally. Instead they are losing water and solute much like in the case of hypoosmotic shock (once the MSCs have opened) and as such have a 'relaxed volume' that is not less than 10% of their initial volume [14]. Future improvements to the collection and processing of data, including the avenues of optimization presented in this thesis, would help to widen this measurement window and provide better insight on the phenomenon of phage-induced volume loss.

5.2 Discussion On the Origin and Measurement Probability of Observed Volume Dips

The loss of volume of a cell in response to phage infection is most likely to be due to the collapse of turgor pressure as water is allowed to escape through the hole. It is beneficial for neither phage nor bacterium for the hole to remain open as in these cases the cell will die. It has been reported that T7 uses a valve comprised of the α10 protein to shut the gp8 tail portal[22]. This valve can open and close, mediating the passage of DNA into the cell and after completion the channel

is left in the closed conformation. The complete genome takes approximately 10 minutes which should lead to the cell becoming completely depressurized. Recalling from equation (4.4), the change in volume per time varies as

$$\frac{dV}{dt} \sim (-\Pi - P), \tag{5.1}$$

and so if a hole is open and permeable to solutes, the volume will decrease. But if a hole is *impermeable* to solutes only P is affected. One expects in this case to see a brief, very small, efflux of water followed by almost immediate recovery as the internal Π is increased (smaller volume, larger internal concentration) to encourage fast flow of water back into the cell. In this solute-impermeable scenario the phage hole essentially becomes another aquaporin, and so even if the hole is open for a long time it is largely irrelevant to the cell as water flux is passive and mediated by solute content/transport. Assuming that the dip is longer than the time that the hole is open for, and noting that the T7 injection time is longer than the measured dip lengths, the hole created by the phage must at some point become impermeable to solutes.

This apparent change in permeability implies there is something fundamentally different between the two data sets I present. Over all infected cells there was a 48%±4% chance of observing a dip, and the likelihood of observing a dip was significantly different between the No Shock and Cold Shock sets. In the control traces there was a (combined) 9%±2% chance of observing a dip, which implies either that there is some natural process which can cause random drops in volume, or the more likely case that the dips in the controls are due to noise only.

5.2.1 The Effect of Noise on Dip Detection

Control Traces

I first assume that there is no time-correlation in the noise found in the cell area traces as the dominant form of noise is segmentation error from DeLTA2. There were traces that demonstrated time-dependent noise which upon review of the videos was generally due to cells moving in the field of view. As such, traces that showed this variant of noise were discarded as if a cell moves or lifts from the slide slightly there is no longer certainty in its area measurement.

I used the No Shock set for noise-modelling as it was the most dip-likely set, and

chose several traces with long division-free sections. I fitted a simple exponential $y = ae^{bx}$, where y is area and x is time to each chosen segment and measured $|y_{fit} - y_{real}|^2$. I then took the square root of the mean of these values to obtain a standard deviation from the fit for that trace. From this, I added Gaussian noise to each point in the fits and generated 100 such 'noise traces', with a random fit being chosen each time for noise addition. In these traces I observed that they had a dip-likelihood of 16%±4%, which is able to explain the dip presence in the No Shock control traces. The fraction of dips observed is given by

$$f' = 1 - (1 - f)(1 - p) = f + p - fp, \tag{5.2}$$

where f is the 'real' fraction of dips and p is the number of false positives due to noise. Therefore

$$f = \frac{f' - p}{1 - p} \tag{5.3}$$

which for the No Shock set is 27%±6% and for the Cold Shock set is 59%±8%. In both cases p is taken as the corresponding control dip count rather than the noise model prediction.

Phage Traces

On the uncertainty in dip duration, it seems likely given the shape of the recovery predicted in Chapter 4 that the true durations of the dips would be slightly less than what is measured. This is because the rate recovery slows down massively when near the "recovered" state (i.e. $V \to V_0$), leading to a shallow V curve. Indeed, when the noise model above is applied to the model data as in figure 5.4, one can see that the dip that would be measured would differ in length from the actual length by an amount of time.

In the case of a non-growing cell this difference is reasonably large, in the modelled case in figure 5.4 the dip loses 700 of its 1700 data-points and the true length is about 40% larger than the measured length. However, for a growing cell the data is skewed upwards as V increases over time. This will act to end the dip early by an amount depending on the growth rate and for dips measured out-with the growth plateau mentioned above this should lead to only a small difference in measured length. This is not expected to change much between the two data sets - however as the Cold Shock set were growing slightly slower, it may have larger underestimation. However, the relationship between dip length and hole-duration

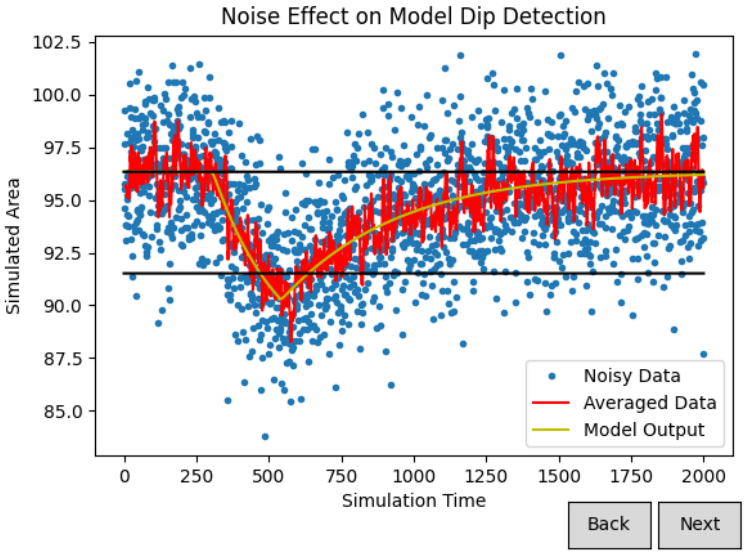


Figure 5.4 *Noise added to the output of the model presented in Chapter 4. The red and blue colours represent data and averaged data as in the analysis above, and the yellow line is model output. The upper black line represents* V_0 *and the lower black line represents* $0.95V_0$ *which is the boundary a dip needs to cross to be measured. One can see the red line crosses the* V_0 *line before the yellow line reaches it, which would artificially shorten the trace. However, this trace does not show cell growth, which would act to significantly steepen the recovery and reduce the degree of shortening.*

is maintained despite this change in proportionality.

5.2.2 Accounting for the Difference Between Cold Cells and Warm Cells

It is clear from these results that there is some property of the bacteria that is temperature dependent that plays an important role in mediating the loss of volume during infection. Broadly speaking, there are two different routes by which solutes and water could exit the cell. One, the channel created by the T7 phage could mediate water and solute loss into the capsid space of the phage. Two, the phage creates a local instability in the membrane through which solutes may escape. The first appears to lie directly against the osmotic model of DNA internalisation presented in Section 1.4.2 as this suggests that the flow of water should initially be in the opposite direction. However, as stated in that section the osmotic model does not have a large amount of supporting evidence as compared to the other models - which are themselves somewhat agnostic to the direction of water flow.

Both of these routes have different potential temperature dependencies, and so this section seeks to explore the effect of temperature in each case and highlight areas of future research to characterise this phenomenon.

Lower Temperatures Cause Cold Shock Protein Up-regulation and Slow T7 Internalisation

Generally bacteria will synthesise new biomass approximately four times faster at 37°C than at 20°C [70]. Between the two data sets presented here those that have been cold-shocked (about 20°C from 37 °C) were approximately 3 times less likely to divide during the experiment, indicating that biomass synthesis was indeed affected. It has been proposed that balanced growth (e.g. exponential) is observed from 20°C to 49°C, however Herendeen has shown that the lower bound may actually be closer to 23°C[45]. Cells in the Cold Shock set were not at room temperature long enough to reach this growth phase (indeed, time on the bench was less than one generation), and so any physiological difference can be ascribed to cold shock. Given this, and that the lower bound for exponential growth may be a few degrees higher than the preparation temperature, it is likely

that the bacteria were cold-shocked and may have begun to up-regulate proteins in response.

Cold-shock in *E. coli* is associated with the up-regulation of Cold-Shock Proteins (CSPs) associated with functions such as translation initiation and DNA supercoiling [56]. The presence of these proteins could in theory interfere with the injection process by binding to the phage DNA. Indeed, in Section 4.3 it is highlighted that the DNA injection process slows down from 60 $bp.s^{-1}$ at 37°C to 10 $bp.s^{-1}$ at 20°C. This would have the effect of increasing the time taken to internalise the DNA to up to an hour. If volume loss was mediated through the phage hole (that is, route one above), this would have the effect of greatly increasing the duration of any dip in the Cold Shock set.

However, this seems somewhat unlikely due to the actual measured durations of the dips. The shortest 'hole-open' time in this scenario is 10 minutes, and from the model proposed in Chapter 4 the dip duration should be much longer than the hole-open time. Extending this to an hour pushes this further out of the realm of possibility. Additionally, this hour-long injection is for cells that are at 20°C. The bacteria in this thesis were infected after about 5-6 minutes of sitting in a small volume at 37°C, and indeed showed lysis times consistent with an injection time similar to the No Shock set. Additionally, if the volume dip corresponds to infection time the lack of significant difference between dip time and lysis time between the sets indicates that there was no significant difference in the infection progress; certainly no difference indicative of the discussed temperature effects.

It is interesting to consider, however, that the growth of a cell leading to a conclusion of a dip is not necessarily actually the real conclusion of that dip. With the value above of 40%, the maximum observable duration would be 12.5 minutes, in the Cold Shock set. At first this appears to allow the volume loss to be DNA-injection-hole mediated, however one must then also pay attention to the relative time periods between hole closing and dip ending. In the model above the hole closes after 240 time-points and has a total duration of 1700 time-points. Therefore in order to have a total duration of 12.5 minutes the hole would have to close after just under 2 minutes. A 10-minute hole, through the DNA-injection hole in T7, would close after 18.6 minutes if this relationship holds. There are experiments recommended in Chapter 7 that would shed some light on the true end of the dip and how the noise affects detection in reality.

Phage in general have their own ways of closing the hole they have used to trans-

locate their genomes. T4, for example, is thought to use a trailing copy of gp2 bound to the genome to seal the hole in the membrane after complete genome internalisation[51]. T7 uses the aforementioned α10 protein helix in gp8. Since the internalisation process takes about 10 minutes, this ought to be the duration of permeability if solutes were escaping through the DNA channel. Within a single set there is at most only a 59% chance of observing a volume dip. This implies that either there is a distribution of hole durations, and therefore volume loss is unlikely to be tied to DNA injection; or that there is a distribution in phage that provoke a loss in volume. There are defective phage particles in every population of phage which may have a different ability to induce volume dips, however they are not particularly well-studied and it is not obvious that their distribution should be dependent on the temperature history of their hosts [36]. Therefore, while they may play a part in the overall visibility of the dips, the effect they have is not something that can be teased out in this work, and is not likely to be the cause of the disparities between data sets.

Lower Temperatures Decrease Membrane Fluidity

The second route is that the phage induces some local instability in the bacterial membrane which is leaky to solutes. Here, the temperature dependence would be in the membrane's ability to 'heal' this instability such that it is no longer leaky. This is a passive process that does not depend on the physiology of the cell, only on the conditions it finds itself in. As the T7 phage injects its DNA through a protein complex this membrane instability would also not affect the infection process, and be independent of the complex being sealed by the phage.

From measurements by Nico Tormeda, unpublished, figure 5.5, it can be seen that a membrane comprised of *E. coli* lipids on a mica sheet undergoes a temperature dependent phase transition from lipid to gel phase. This transition is quick to begin when the ambient temperature is changed and appears to begin as a nucleation process but proceeds spinodally over a long period of time. This s supported by existing reports that the actual bacterial membrane does undergo partial phase change to a gel phase when the cells are exposed to lower temperatures [42]. The work by Tormeda shows that the recovery of the membrane to a liquid phase may not be fast enough to have been complete by the time phage are introduced.

Proteins in the membrane will act to raise the temperature at which this occurs

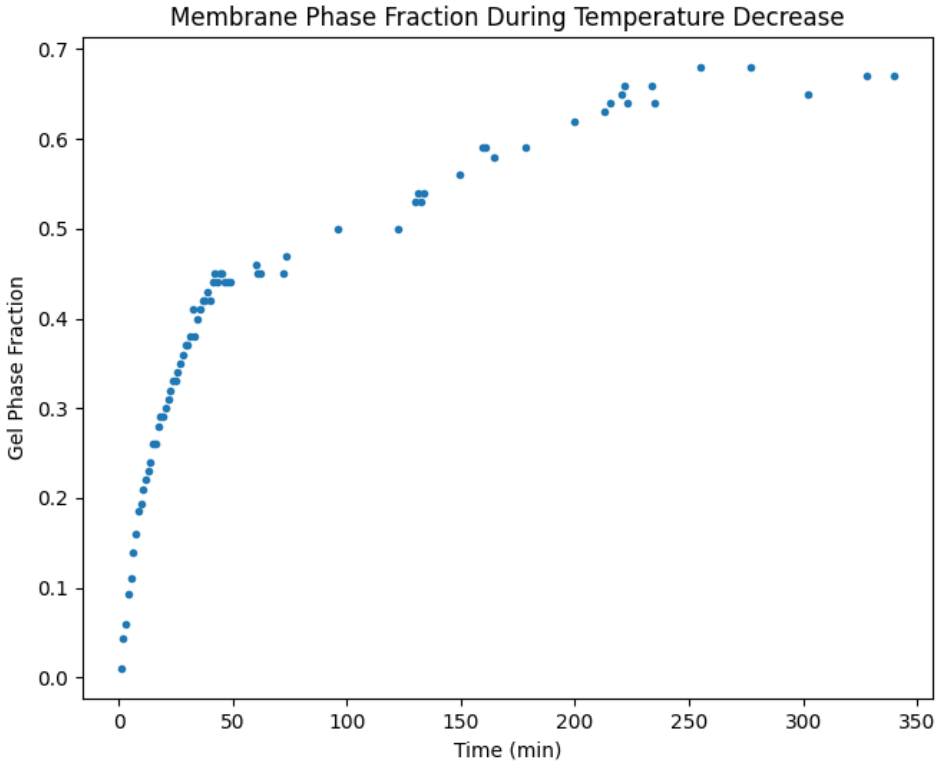


Figure 5.5 *Experimental data gathered by Nicolo Tormena (University of Durham/University of Edinburgh) showing membrane gel phase fraction as a function of time. The temperature was 40°C at $t = 0$ and subsequently lowered to 23°C at a rate of 1°C per second. The phase was measured by AFM, which limited the rate of cooling due to increased thermal noise. This experiment was performed on a muscovite P1 supported lipid bilayer of 3:1 POPE-POPG (Avanti Polar Lipids) diluted in chloroform.*

by imposing local order on the lipids. Additionally, the presence of Bayer patches (Section 1.3.1) may also raise this temperature by the same mechanism, reducing the freedom of movement of lipids. If one considers the membrane around the injection site unstable and leaky, it is reasonable to assume that if the lipids are less mobile the instability will be leaky for longer. However, this is proposed using data from a lipid bilayer on a mica sheet, the actual physical behaviour may be quite different. One may also expect this to have an effect on the dynamics of the osmotic shock response.

5.3 Concluding Remarks and Future Direction: Label-Free Infection Measurement

In this chapter an observable reduction of volume in growing *E. coli* cells due to phage infection is reported. The phenomenon appears to have a reasonably strong dependence on the temperature history of the cell - and likely also on the measurement temperature. Several potential causes of the volume dips have been discussed, but more work needs to be done before any can be explicitly ruled out.

Firstly, it is not explicitly known that this effect is due to a single phage infecting a bacterium, it could be the combined effect of multiple. It is also not explicitly known if the volume dip corresponds with the exact point of DNA injection initiation. These things are assumed, but the work has not been done to confirm them. The usage of labelled phage and nuclei-binding dyes would allow the time-correspondence of infection, phage number, and volume dip to be uncovered. The method of observing the volume dip is expected to be agnostic of any label as the neural network analysis pipeline does not need fluorescent cells. Therefore, the loss of volume is a *label-free* marker of phage infection, and so finding the correlation between time of DNA injection and volume dip and how they interact with phage number is of a high importance to develop the technique.

The dependence on temperature of this volume loss may also allow, if the hypothesis of a membrane instability is verified, the exploration of different bacterial membrane states via directed volume loss. To explore this further, more experiments at different temperature points and at different times between cooling and warm-imaging would be highly valuable, with the work provided by Tormeda giving a curve of time-of-exposure that could be followed. The model provided in Chapter 4 may be improved by working in the temperature dependence in some way, allowing the effects of cold shock gel-formation to be captured mathematically.

Chapter 6

Modelling the Infection Dynamics of a Single *E. coli* Cell by Phage T4

During the second year of my program the labs were closed for an extended period as the country was put into covid-19-induced lockdown. Faced with an uncertain number of months at home without access to lab facilities it was decided that it would be beneficial to have a mathematical model of the phage infection process. At the time we believed we were working with T4, and so this was the phage chosen to model. There existed a model written by Weiße *et al* [138] and later improved by Thomas *et al* [130] which described the general growth state of an exponentially growing cell with regards to transcription, translation, and DNA replication. As these are all vital processes that are inhibited or hijacked by T4, this model provided a good base for a phage-adapted version. Generally speaking, with detail to follow, the T4 phage infection (and to a similar extent the T7 infection) is temporally controlled via transcription and the expression of self-regulating proteins. Therefore while the source of the dips in Chapters 4 and 5 is unknown, a model that takes protein expression explicitly into account could be used to predict whether or not the dips are physical (membrane instability) or biological (cold shock induced biosynthesis slow-down) in origin. Additionally, by changing certain parameters in the model it is possible to simulate the presence of different antibiotics, further showing how different conditions affect the phage infection.

Due to time constraints and the rush back to physical laboratory work post-lockdown, this model exists only in mathematics. Tables of reactions and

parameters are provided, though no code is available.

6.1 Background Theory of the T4 Host Takeover

In general for a complete lytic infection the virus must completely overtake the host's biomachinery to ensure maximum resource availability for virion production. The exact processes by which this occurs is generally unique to the phage, and can comprise some quite complex patterns of DNA transcription and subsequent protein translation. Compared to other phages, there is an abundance of information regarding the T4 infection. However, like every phage, there is much that is still unknown and many genes within the T4 genome have no known function. While I did not anticipate that the model would be able to ascertain protein function specifically, it may be able to predict how certain processes of the infection are affected by different antibiotics, for example. In this way, it may be used both as a tool to highlight future experiments as there are parameters that require fitting to experimental data and as a predictor for the change in infection dynamics to extracellular influences. Before delving into the base model and the adaptations made, I will provide some information regarding the infection process of *E. coli* by T4.

An overview of the steps before the processing of T4 DNA within the cell is found in Section 1.3.2, in the model I consider those processes to be instantaneous in time and so only describe the transcription, translation, and DNA replication of T4 during infection. Similar to T7, T4 engages in three transcriptive periods: 'early', 'middle', and 'late'; as seen in figures 6.1 and 6.2. I here define the 'early period' as the time period wherein transcription of genes under the 'early' promoters is dominant, and so on for the middle and late periods. T4 differs from T7 in the methods of temporally switching periods. T4 has a large genome of approximately 170 kbp encoding about 300 gene products [85]. The early transcriptive period encodes specific proteins that encourage host RNAP enzymes to transcribe middle promoters, and so on for middle to late. Compared to the feedback-loop based temporal switching in T7, wherein an expressed protein will inhibit its own expression[114], T4 takes a more active route of expressing specific proteins to inhibit earlier products.

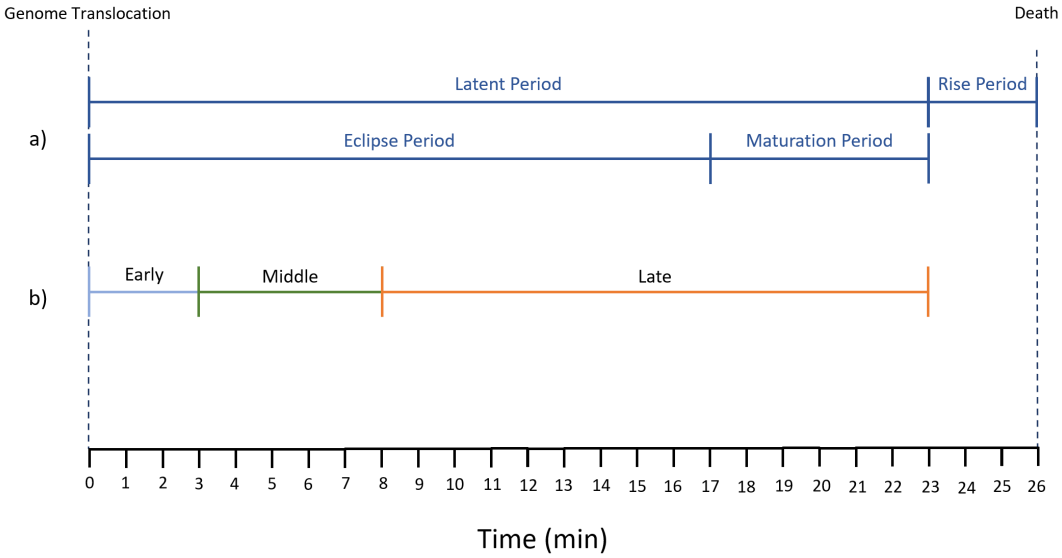


Figure 6.1 *Rough timeline of the transcriptive periods of T4 against the periods of phage development.* ***a)*** *phases of phage development. It is less important that these are exactly correct, and are more of an estimate based on the literature [41]. Latent period is defined as the time between genome translocation (treated as the start of the infection) and the beginning of the rise period. Rise period is defined as the time over which an otherwise simultaneously infected population of cells will lyse (i.e. one 'step' increase in phage numbers - for a single cell this period is instantaneous). The Eclipse period is the time between genome translocation and the maturation of the first phage particle. The Maturation period is defined thus as Latent minus Eclipse and is the time over which a cohort of virus particles matures.* ***b)*** *the transcriptive periods in a single cell. Early takes place over the first 1-3 min, middle (and delayed early) from 3-8 min, and late from 8 min on-wards. No middle or early genes are expressed after approximately 15 min.*

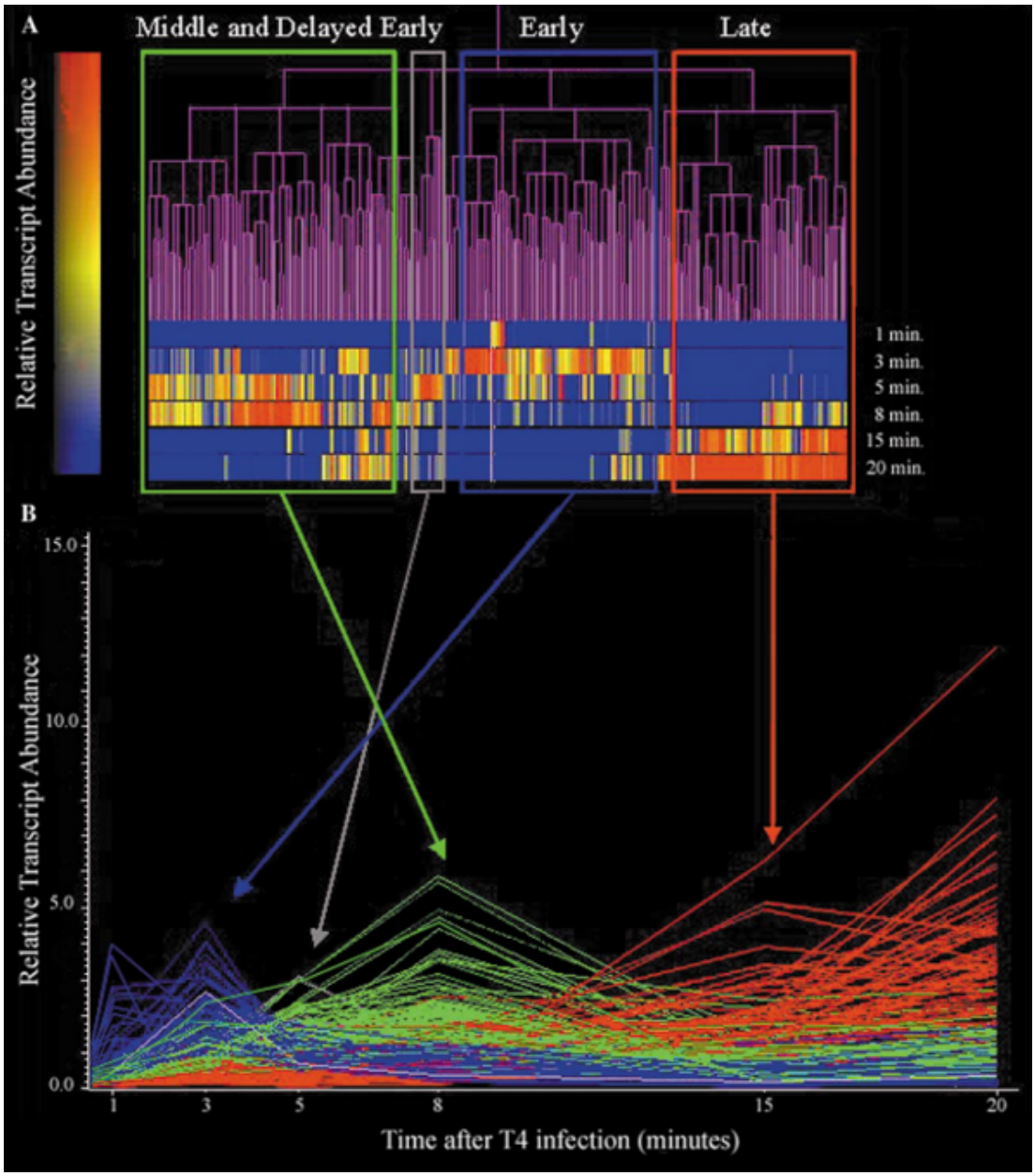


Figure 6.2 *The relative abundance of mRNAs associated with Early (blue), Middle (green), and Late (red) promoters during a T4 infection. This data was obtained by Luke et al who performed bulk microarray measurements of E. coli B. For the purposes of the model I do not consider Delayed Early (grey) to be its own distinct transcriptive period. Inset displays relative expression of each separate gene, with blue as 0 and red as 1. Image taken from [78].*

6.1.1 Early Transcription

The early period is the most complex, as the phage needs to quickly engage in shutting down host processes that may interfere with the infection and begin switching bacterial expression to phage as quickly as possible. When the DNA is internalised, there are approximately 20 times the number of host promoters per DNA molecule as early T4 promoters (quantitatively, 39 phage to 650 host[99][118]) that compete for the RNAP holoenzymes in the exponentially growing cell. Yet despite this, the transcription of early gene products begins immediately after infection[47].

Primary in this process is the fact that T4 early promoters contain excellent matches to the σ^{70}-associated RNAP elements. This is directly associated with a disproportionately higher 'strength' of promoter that means binding events between RNAP and early promoters are much more successful than those between RNAP and host promoters. T4 early promoters have been described as 'über' promoters for this strength, and for the fact that they appear to contain other pieces of information that optimise the interaction further between host RNAP and the promoter elements [47].

Further to this is a capsid protein called Alt, which enters the cell alongside the DNA payload. The Alt protein is a mono-ADP-ribosyltransferase, meaning it catalyses the transfer of the ADP-ribosyl moiety onto the arginine or histidine residues of a target protein. For context, ADP-ribosylation is a reversible post-translational process which is usually involvd in key cellular processes such as DNA repair, cell signalling, and apoptosis. In this case it ADP-ribosylates the Arg265 residue on one of the two α subunits of RNAP [26]. It also modifies some host proteins including other RNAP subunits and those involved in translation and metabolism. The Alt target, α-Arg265, has been demonstrated to be crucial for host transcription, in the interaction of the C-terminal of host RNAP (α-CTD) with the ribosomal promoter *rrn1 B1*-associated UP (upstream) element [112]. This UP element stimulates transcription of ribosomal RNA, however overexpression of Alt in the absence of phage does not appear to inhibit cell growth, nor does it appear to impair normal cell transcription. Instead this modification to host RNAP appears to stimulate particular early promoters leading to a two-fold enhancement in their activity [140].

Another strategy by which T4 may assert transcriptional dominance in the early

period is the activity of the Alc protein, which is expressed a little after the initial burst of early transcription. The purpose of Alc is to direct actively-transcribing RNAP molecules to frequent terminations, effectively shutting down host transcription once a sufficient Alc concentration has been reached. It recognises DNA that contains unmodified cytosine and prevents transcription of that DNA. Host DNA is unmodified, whereas T4 DNA contains hmCyt - which is hydroxymethylated cytosine - and so the shutoff of host transcription in this manner does not affect T4 transcription [59].

A short time after the infection two new ADP-ribosyltransferases are expressed, modA and modB. Like Alt, modA ADP-ribosylates the α subunits of RNAP, however targets both subunits rather than just one. The consequence of this is that the RNAP can no longer recognise the UP element and thus has its activity reduced significantly. Interestingly, the deletion of modA does not slow the rate of decline in host transcription, suggesting that this is a secondary function of modA [26]. ModB targets host translation factors (necessary for initiation of translation[147]), the 30S subunit of the host ribosome[64], and trigger factor (a ribosomal chaperone to aid protein folding[48]). This is a relatively comprehensive inhibition of host translational capability and as such allows T4 to effectively redirect all ribosomal activity towards its own genome.

6.1.2 Middle Transcription and Early-Middle-Late Transcriptional Switching

The T4 middle genes primarily comprise proteins for replication, recombination, and nucleotide metabolism. By the start of the middle period there is little host activity (both transcription and translation) remaining. The end of the middle period is the time when phage DNA replication can first begin with the expression of the sliding clamp protein gp45, which both forms part of the phage DNA polymerase and enhances transcription by interacting with the T4-encoded sigma factor gp55[119].

As the infection progresses T4 begins to express RegB, an endoribonuclease which causes a fast degradation of early mRNAs. Additionally, the host DNA template is removed from host RNAP enzymes, presenting an effective hard-stop for host transcriptional activity. T4-encoded nucleases (EndoII and IV) selectively degrade dC-containing DNA, so that after the first few minutes of infection there

is essentially no host DNA to transcribe by existing RNAP molecules. Early transcription is halted by expression of two early gene products, AsiA and MotA. The former interacts with portions of σ^{70} in order to inhibit recognition of early T4 transcripts (inhibiting early transcription), and at the same time stimulates transcription of middle T4 transcripts. The interaction between AsiA and σ^{70} is thought to be thedriving force behind the transition between early and middle transcription. The latter protein, MotA (different from the motA of the host flagellar motor system), is a DNA-binding transcriptional activtor facilitating middle promoter recognition and activation.

The switch to late transcription is facilitated through more ADP-ribolsylation of the RNAP α subunits and binding of the phage encoded RpbA protein to the RNAP core allowing late promoters to out-compete middle promoters. Further, there is a replacement for σ^{70}, gp55 - the smallest known sigma factor, that binds to RNAP to facilitate recognition of late promoters. The efficiency of gp55 is enhanced significantly by a coactivator protein gp33 CITE.

6.1.3 Late Transcription

T4 late transcription is dedicated to the expression of phage structural proteins (capsid, tail e.t.c.), several virus assembly factors, and the proteins responsible for cell lysis. The eight tRNAs that T4 encodes are all found clustered in the late transcriptome, however early and middle promoters also direct transcription into the cluster [85]. These tRNAs may be a recent acquisition for the phage, and enhance expression of some of the lower-abundance late gene products.

6.1.4 Translation

The switch from host to phage translation is swift. Nearly no host proteins are detectable by 2D gel electrophoresis after 1 minute of infection (that is, there is virtually no host translation after 1 min from DNA internalisation) [85]. It is reported that T4 modifies the translation apparatus of the cell, and synthesizes new ribosome-binding proteins; however most of the genes responsible for this have yet to be properly identified.

6.1.5 Lysis

Important for the model and vitally important for the phage infection is the lysis event wherein a cell containing approximately 100-150 T4 phage particles bursts to release the progeny into the environment. Two proteins are involved in this process: gpe and gpt. Gpe is the T4 lysozyme, which breaks down the peptidoglycan later after passing through a hole created by gpt - the T4 holin protein - in the inner membrane [57].

Rather interestingly, gpt (or T4 lysozyme, or T4L) is not strictly late, as mRNAs containing the sequence for it are produced in the early period. However these early expressed gpts appear to form very stable secondary 'hairpin' structures which encompasses both the AGGAGG sequence that the ribosomes recognise and the AUT start codon. Thus these early transcripts are rendered completely un-translatable and only the non-hairpin transcripts produced in the late period are translated. When early transcripts are forcibly translated by unwinding the hairpin structure, the rate of translation is aproximately 100 times less than that of the late transcripts and so their presence is somewhat of a mystery. The late gpt transcripts themselves begin to be produced when phage DNA replication begins, whereupon two late promoters are activated.

6.2 Modelling the Death of a Growing Bacterium

6.2.1 Foundation Model Reactions and Equations

I built my model of phage infection upon an existing description of a single *E. coli* cell growing exponentially, wherein gene expression and cell growth are coupled under three broad trade-offs faced by all cells in nature: finite energy, finite resources, and a finite proteome. The foundation model is able to accurately report changes in growth rate as a function of extracellular nutrient and the intracellular protein:mRNA ratio - which has been shown experimentally to be linearly dependent on the growth rate (see fig 1a in). As it is capable of responding to the presence of certain antibiotics accurately, it presents a measure of the response of the cell to extracellular perturbation. During the initial preparation of the model I considered that it was helpful to observe responses in infection processes to differing nutrient levels to predict environmental effects

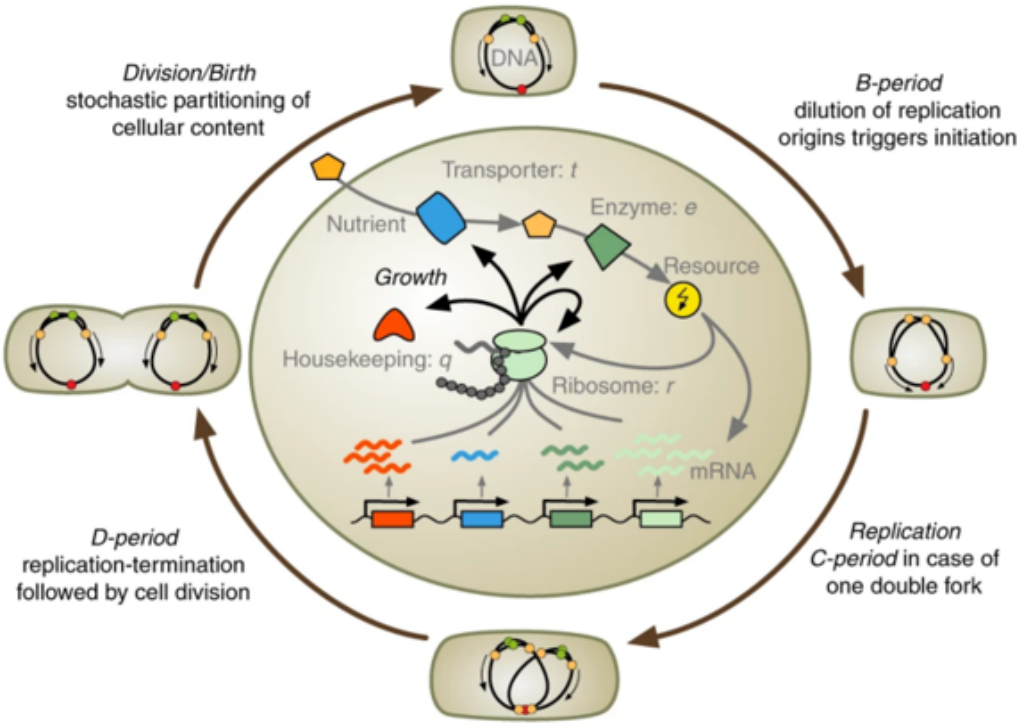


Figure 6.3 *Schematic cartoon of the foundational model by Thomas et al. Nutrient is imported by transporters and catalysed into energy by enzymes. Energy is then used in transcription and translation processes. Each species is diluted at the growth rate in order to retrieve steady state results. The DNA cycle allows the description to be specifically single-cell by tracking mass as a function of cell divisions which are dependent on the number of replication origins (red dots) and forks (yellow dots) as in Donachie CITE. Image taken from [130].*

on phage growth rate.

A schematic of the foundation model is seen in figure 6.3. The model is built on a set of reactions that govern the import and catalysis of energy and its subsequent use in the various intracellular processes. I have collected these reactions as seen in table 6.1. These reactions give rise to a list of first order ordinary differential equations, through which the growing cell is modeled. It is important to note that in the case of both models the canonical species name is used also for the functional variable, i.e. the reaction species m_r will be referred to in the numerical equations as m_r. All numerical equations inherently deal in concentrations of molecules.

Firstly, the import of nutrients to the cell and their subsequent catalysis to energy is driven by enzymes. The time-dependent ODE for intracellular nutrient s_i is

$$\frac{ds_i}{dt} = \nu_{imp}(e_t, s) - \nu_{cat}(e_m, s) - \lambda s_i, \tag{6.1}$$

where e_t and e_m are the numbers of transport and catalysis enzymes respectively. The rates ν_{imp} and ν_{cat} are assumed to have Michaelis-Menten kinetics and thus follow

$$\nu_{imp} = e_t \frac{v_t s}{K_t + s}, \nu_{cat} = e_m \frac{v_m s_i}{K_m + s_i}, \tag{6.2}$$

where v_t and v_m are the maximum rates and K_t and K_m are the half maximal thresholds of s and s_i. The amount of intracellular nutrient is diluted by the growth rate λ due to the cell increasing in size as it grows.

The cell energy a is governed by the equation

$$\frac{da}{dt} = n_s \nu_{cat}(e_m, s_i) - \sum_{x \in \{r,t,m,q\}} \Big(\nu_x(c_x, a) - k_b r m_x + k_u c_x \Big) - \lambda a, \tag{6.3}$$

where the sum accounts for energy that is involved in the creation, degradation, and completion of transcription complexes c_x (using one a molecule per complex). The complexes are formed of one ribosome r and one mRNA molecule m_x. The three ODEs governing these are, where $x \in \{r, t, m, q\}$,

$$\frac{dm_x}{dt} = \omega_x(a) - (\lambda + d_m) m_x + \nu_x(c_x, a) - k_b r m_x + k_u c_x, \tag{6.4}$$

$$\frac{dr}{dt} = \nu_r(c_r, a) - \lambda r + \sum_{x \in \{r,t,m,q\}} \Big(\nu_x(c_x, a) - k_b r m_x + k_u c_x \Big), \tag{6.5}$$

and

$$\frac{dc_x}{dt} = k_b r m_x - k_u c_x - \nu_x(c_x, a) - \lambda c_x. \tag{6.6}$$

The proteome in the modelled cell consists of four species, denoted by the above set $x \in \{r, t, m, q\}$ (as defined above in table 6.1). The equations for the enzymes and housekeeping proteins are

$$\frac{de_t}{dt} = \nu_t(c_t, a) - \lambda e_t, \tag{6.7}$$

$$\frac{de_m}{dt} = \nu_m(c_m, a) - \lambda e_m, \tag{6.8}$$

and

$$\frac{dq}{dt} = \nu_q(c_q, a) - \lambda q. \tag{6.9}$$

Transcription and translation are derived from Michaelis-Menten kinetics with energy as the substrate, where the transcription of new mRNAs follows

$$\omega_x(a) = \frac{w_x a}{\theta_x + a}, \tag{6.10}$$

and the translation of those mRNAs to proteins follows

$$\nu_x(c_x, a) = \frac{\gamma(a)}{n_x} c_x \tag{6.11}$$

with an effective translational elongation rate of

$$\gamma(a) = \frac{\gamma_{max} a}{K_\gamma + a}, \tag{6.12}$$

where K_γ is the half-maximal threshold of a for $\gamma(a)$.

6.2.2 Cell Growth

The Relationship Between DNA Initiation and Cell Division

Before describing explicitly the form of λ used in the model, I will provide a short overview of work by Donachie [28] wherein he presents that there is a fixed ratio between mass and number of DNA replication origins at time of division. An observation by Cooper and Helmstetter is that there is a fixed time, denoted as τ_{C+D} between the initiation of DNA replication and subsequent cell division (approximately 60 min for *E. coli* B/r)[20]. This period is longer than the doubling time of *E. coli*, and so an exponentially growing cell will naturally contain multiple replicating strands of DNA.

An exponentially growing cell will divide when its mass has doubled, so one may say that the mass at replication initiation is a multiple of a fixed mass for each cell at different growth rates. This relationship is shown graphically in figure 6.4. The mass at initiation can thus be written as

$$M_i = \begin{cases} M & \text{for } \lambda \leq 1 \\ nM & \text{for } 1 < \lambda \leq 2 \\ 2nm & \text{for } 2 < \lambda \leq 3 \end{cases} \tag{6.13}$$

where M is that fixed mass and $n \in Z$ is the the number of doublings per τ_{C+D}

	Transcription	Degradation	Ribosome Binding	Degradation
Ribosomes r	$\emptyset \xrightarrow{\omega_r} m_r$	$m_r \xrightarrow{d_m} \emptyset$	$r + m_r \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_r$	$n_r a + c_r \xrightarrow{\nu_r} r + m_r + r$
Transporter Enzyme e_t	$\emptyset \xrightarrow{\omega_t} m_t$	$m_t \xrightarrow{d_m} \emptyset$	$r + m_t \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_t$	$n_t a + c_t \xrightarrow{\nu_t} r + m_t + e_t$
Metabolic Enzyme e_m	$\emptyset \xrightarrow{\omega_m} m_m$	$m_m \xrightarrow{d_m} \emptyset$	$r + m_m \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_m$	$n_m a + c_m \xrightarrow{\nu_m} r + m_m + e_m$
Growth-Independent Proteins q	$\emptyset \xrightarrow{\omega_q} m_q$	$m_q \xrightarrow{d_m} \emptyset$	$r + m_q \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_q$	$n_q a + c_q \xrightarrow{\nu_q} r + m_q + q$

Table 6.1 *Table showing the set of reactions that govern the foundational model by Thomas et al. mRNAs, m_x, are generated from an infinite pool at their respective transcription rates ω_x, and spontaneously degrade at rates d_x. Ribosomes bind to free mRNAs to form complexes c_x in reversible reactions (the two proteins may spontaneously unbind) at rate k_b. The complexes will then degrade into ribosomes, free mRNAs, and the protein product of the mRNA.*

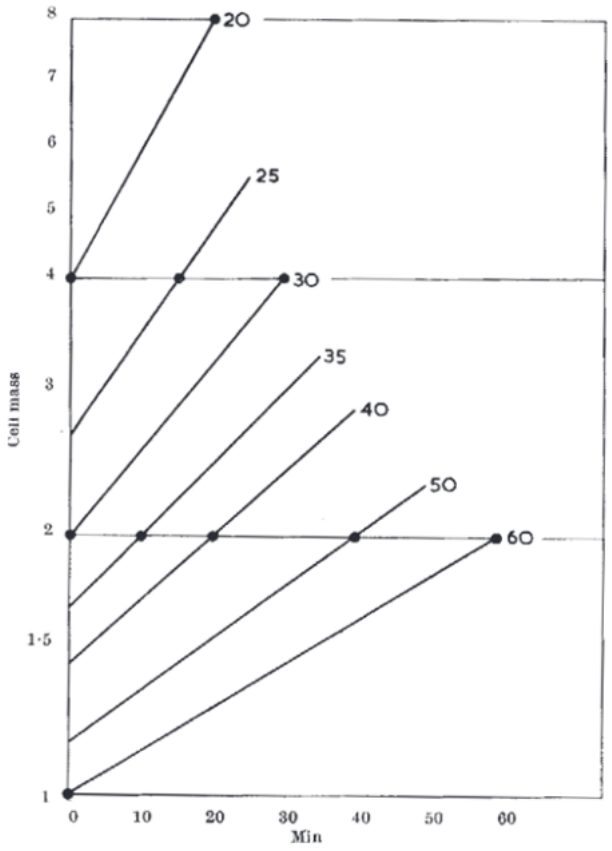


Figure 6.4 *Increase in cell mass of cells at different growth rates (doubling time indicated at end of each line). The cells are assumed to grow across one cell-cycle beginning with the end of a division event at 0 min. Because there is a fixed period of 60 min between initiation and division it is possible to calculate the time at which initiation occurs for a cell growing at a given rate. These initiation times are marked as black circles on the horizontal lines. Figure taken from [28].*

(hr^{-1}). This relationship is shown graphically by Donachie. The constant time period τ_{C+D} is not necessarily greater than the doubling time T, and so the number of replication origins per cell will change with the growth rate as

$$N_i = \begin{cases} 1 & \text{for } \lambda \leq 1 \\ 2 & \text{for } 1 < \lambda < 2 \\ 3 & \text{for } 2 < \lambda < 3 \end{cases} \tag{6.14}$$

where N_i is the number of origins in a cell at initiation. Both N_i and M_i vary discontinuously with the growth rate λ (hr^{-1}), and both change in the same way such that

$$\frac{M_i}{N_i} = constant, \tag{6.15}$$

which holds for cells of any size growing at any rate. The consequence relevant for this model is that, given the rate at which mass increases in an exponentially growing cell,

$$\frac{dM}{dt} = \lambda M = \frac{M}{T}, \tag{6.16}$$

where T is the doubling time. The mass of the cell at any point in time is given by

$$M_t = M_0 e^{\frac{t}{T}}, \tag{6.17}$$

and thus the mass at division follows

$$M_d = \frac{M_i}{N_i} e^{\frac{C+D}{T}}, \tag{6.18}$$

where $t = (C + D)$ since division occurs after this constant time, at a mass $M_0 = \frac{M_i}{N_i}$. Hence the log of mass at division varies as

$$\ln M_d = ln \frac{M_i}{N_i} + \frac{C + D}{T}, \tag{6.19}$$

and is thus proportional to $\frac{1}{T}$ provided the mass at initiation is constant.

Modelling Cellular Growth in the Deterministic Limit

Now with the mass function from Donachie in hand, one turns to the model by Thomas *et al*, which comprises a set of coupled Langevin equations that can be found in [130]. These equations were built to account for various noise sources in

growth, and as I take the noise-free deterministic limit the terms I recover are

$$\frac{d\bar{\mathbf{X}}}{dt} = \nu\mathbf{f}(\bar{\mathbf{X}}) - \lambda(\bar{\mathbf{X}})\bar{\mathbf{X}}, \tag{6.20}$$

and

$$\frac{1}{\bar{M}}\frac{d\bar{M}}{dt} = \lambda(\bar{\mathbf{X}}) - \frac{1}{2}dD(t), \tag{6.21}$$

where $\bar{\mathbf{X}}$ are the mean concentrations of each molecular species, ν is the stoichiometric matrix of parameters for the reaction rate equations $\mathbf{f}(\bar{\mathbf{X}})$. $D(t)$ is a process that counts the divisions of the cell and $\bar{M}$ is the mean cell mass. Since the dynamics of the mean concentrations are independent of mass and divisions, the concentrations reach a steady state given by

$$\nu\mathbf{f}(\bar{\mathbf{X}}) = \lambda(\bar{\mathbf{X}})\bar{\mathbf{X}}, \tag{6.22}$$

and the mean cell mass increases as

$$\bar{M}(t) = M_0 e^{\lambda(\bar{\mathbf{X}})t}. \tag{6.23}$$

λ is the growth rate such that cell divisions occur once every $\tau = \frac{ln2}{\lambda}$, so the mass at cell birth can be obtained from the timing of initiation,

$$M_0 = \frac{e^{\lambda(\bar{\mathbf{X}})\tau_{C+D}}}{2O_c}, \tag{6.24}$$

which is analogous to Donachie in equation 6.18. Here O_c is the critical concentration of origins that occurs at initiation given by 2^F where F is the number of fork generations. In the original model the division counting process $D(t)$ is Poissonian, and thus in order for the model to avoid probability and noise it must be made deterministic as well. If an initiation is always followed by a division, one might expect the number of initations to be given by

$$I(t) = \left\lfloor \frac{t}{\tau} \right\rfloor, \tag{6.25}$$

and thus the number of divisions by

$$D(t) = I(t - \tau_{C+D}) = \left\lfloor \frac{t - \tau_{C+D}}{\tau} \right\rfloor, \tag{6.26}$$

such that division has only occurred after a constant period after initiation has occurred. In this way it does not depend on O_c, however this dependence could

be inserted by redefining $D(t)$ to contain the Heaviside function

$$H\Big(O_c - \frac{2^F}{M}\Big) = \begin{cases} 0 & \text{for } 2^F/M > O_c \\ 1 & \text{for } 2^F/M \leq O_c \end{cases}. \tag{6.27}$$

While the exact mechanism of control over cell division remains a somewhat open question, this model has been used here as the answer for our phage model.

6.2.3 Adapting the Foundation Model for Phage Infection

The phage-adapted model (henceforth phage model) is split into three coarse boxes: Phage Clock, Phage Structural, and Phage Replisome. These boxes include what has so far been identified as key information for the progression of the infection. A schematic of the phage model can be seen in figure 6.5.

Phage Clock

The Phage Clock includes proteins that are responsible for the transcriptional control of host biosynthesis machinery and for the progression of the infection cycle through the periods listed in figure 6.1. As such, phage mRNAs are contained in this block, and it is the largest of the three. I define three vectors of mRNAs $\overrightarrow{m_E}$, $\overrightarrow{m_M}$, and $\overrightarrow{m_L}$ for early, middle, and late transcription products. These vectors contain the concentrations of mRNAs specific to those transcription periods as discussed in Section 6.1. These are proportional to their respective transcription rates ω_p, and the concentration of RNAP molecules with affinities for each period's promoter. It is not obvious that each rate should be different, and so in the first instance each phage product is assigned the same constant rate. Some phage products have their transcription mediated by others, and so in those cases the specific transcription rate is used.

It was decided to have a species of RNAP for each promoter as T4 uses host core RNAP and modifies it with various proteins. In principle one may take these modifications to be creating a 'new' RNAP molecule since its promoter affinity is different to the previous modification. During the early period host RNAP is mostly unmodified and is still capable of transcribing host DNA, and so is named R_E. During the middle transcription period MotA and AsiA bind to R_E to form

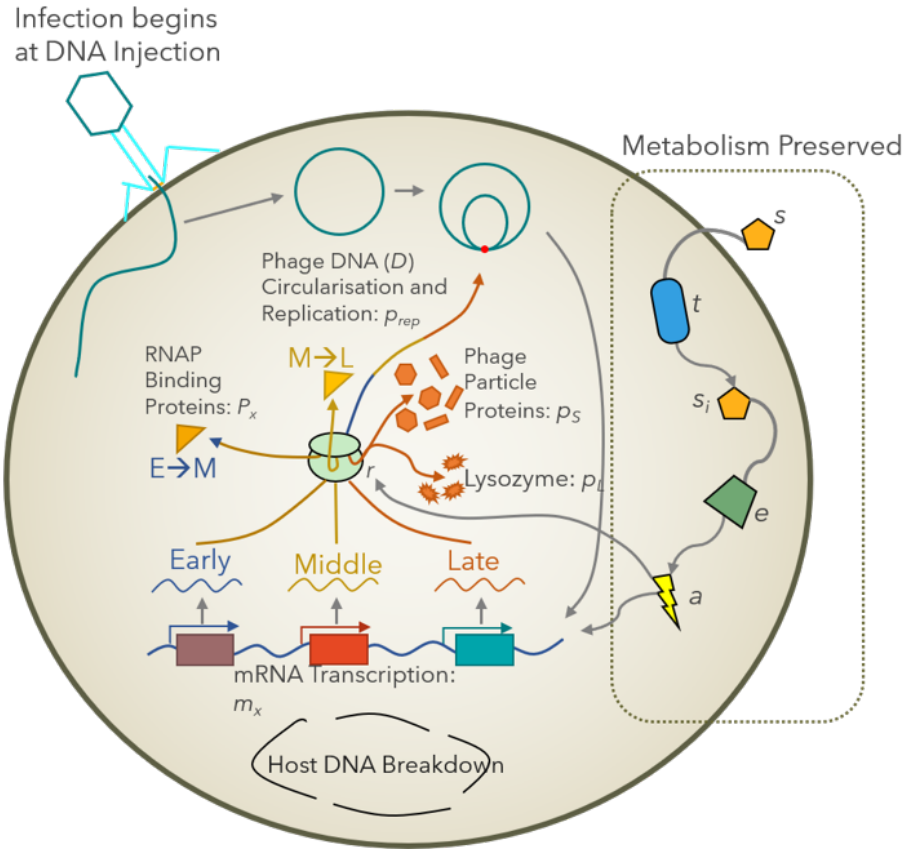


Figure 6.5 *Cartoon schematic of the phage-adapted model. The metabolism that catalyses imported nutrients is maintained, but the transcription and translation processes are altered from host to phage. Transcription occurs in three distinct periods defined by the expression of proteins that reconfigure host RNAP molecules to have affinities for the different promoters as described in the Phage Clock.*

R_M. The switch to late transcription gp55 will replace σ^{70} and gp33 will bind to R_M to form R_L. Another middle gene product gp45 is required for both DNA replication and late transcription initiation, and so this protein p_{45} will regulate $\overrightarrow{w_L}$.

For all modifications of RNAP it is assumed that it binds irreversibly to DNA and that the rate is effectively infinite. Thus, if an unmodified RNAP molecule exists at the same time as a protein that would modify it it will be modified immediately. This assumption may not hold, in which case a proper binding reaction can be defined later.

During the early period I define a set of transcripts (mRNAs) as

$$\overrightarrow{m_E} = [m_{alc}, m_{reg}, m_{em}, m_{end}, m_{rep}], \tag{6.28}$$

where the subscripts are as follows: *alc* is the Alc protein; *em* is a combination (representing both) of MotA and AsiA; *rep* are the DNA replisome proteins; *reg* is the RegB protein; and *end* are EndoII and EndoIV. Synthesis of these proteins proceeds as in the foundation model following Michaelis-Menten kinetics. Alc simply has the effect of reducing host transcription ω_x by an amount proportional to its number. The *em* proteins bind to R_E to form R_M. RegB is a negative influence on the number of all $\overrightarrow{m_E}$, including itself. The *end* proteins will reduce the host DNA number to effectively zero - host DNA is not explicitly modeled and so the effect of these proteins in the model will instead be to inhibit $\overrightarrow{w_x}$ alongside Alc.

Switching from the early to middle periods depends on the following reactions for the production of the *em* proteins and subsequent binding to R_E. The synthesis of *em* follows

$$m_{em} + r \underset{k_b}{\overset{k_u}{\rightleftharpoons}} c_{em} \xrightarrow{v_x} p_{em} + r + m_{em}, \tag{6.29}$$

where the binding of a ribosome to an mRNA molecule to form a complex is reversible. The formation of R_M follows

$$p_{em} + R_E \xrightarrow{\infty} R_M, \tag{6.30}$$

where the infinite rate denotes that as soon as a p_{em} exists, it will bind to an R_E. The transcription rate w_p is unaffected by the number of R_M, however the overall rate of production of middle transcripts $\overrightarrow{m_M}$ is as transcription complexes b_M are formed between R_M and T4 DNA D_{T4}. This formation of transcription

complexes is the same for all R_x and b_x where $x \in \{E,\ M,\ L\}$ and follows

$$R_x + D_{T4} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} b_x \xrightarrow{w_p} \overrightarrow{m_x} + D_{T4} + R_x. \tag{6.31}$$

For the middle period I define the following set of transcripts:

$$\overrightarrow{m_M} = [m_{45}, m_{ml}, m_{rep}], \tag{6.32}$$

where 45 is gp45, the sliding clamp protein; *ml* is a combination of gp33 and gp55, and *rep* are the proteins required for DNA replication, the synthesis of which is maintained throughout the early and middle periods. The synthesis of middle proteins follows the form of equation 6.29 and the switching of middle to late transcription follows equation 6.30.

The key late gene products are those that are responsible for the structure of phage and the lysis of the cell. Temporal control is swapped from re-configuration of RNAP to expression of holin proteins and so I define for the late period

$$\overrightarrow{m_L} = [m_S, m_{ly}, m_{ho}], \tag{6.33}$$

where S denotes structural proteins; ly denotes lysozyme; and ho the holin proteins. It is assumed that the rate of assembly of new phage is instantaneous and that the assembly proteins are contained within S. A combination of lysozyme and holins are required for lysis but the timing is dependant on the holins as they regulate access to the cell wall. Thus one can express the probability of lysis occurring as

$$P(Lysis) \propto l_N H(h), \tag{6.34}$$

where l_N is the normalised amount of lysozyme and $H(h)$ is the Heaviside function

$$H(h) = \begin{cases} 1 & \text{if } h > 0 \\ 0 & \text{if } h < 0 \end{cases}. \tag{6.35}$$

Phage Structural

The assembly of new phages takes place during the late phase, and the rate of maturation determines how many progeny an infection can produce. The number

of current virions per bacterium is modeled as

$$N_{ph} = \left\lfloor \frac{p_s}{n_{ps}} \right\rfloor, \tag{6.36}$$

where N_{ph} is the number of phage particles and n_{ps} is the number of structural proteins p_s that fit into one phage. The division is within a floor function so that $N_{ph} \in Z$. The burst size of the phage is therefore dependent on how much p_s can be made before lysis.

Phage Replisome

Replication of phage DNA begins in the early period and maintains a steady rate until the middle period when gp45 is synthesized, at which point the rate increases dramatically [66]. T4 DNA polymerase elongates at a rate 10 times that of RNAP and can pass over a currently elongating RNAP molecule without slowing down. Thus the formation of transcription complexes does not reduce the pool of DNA molecules. The exact process of DNA replication is difficult to model as there are 7 replication origins on the genome which activate at different times and are dependent not only on R-loop[43] formation but also recombination proteins. Thus a simplified form of replication complex D^* and new DNA molecules follows the reactions

$$p_{rep} + D_{T4} \underset{k_r}{\overset{k_{-r}}{\rightleftharpoons}} D^* + D_{T4}, \tag{6.37}$$

where D_{T4} is the concentration of T4 DNA, and

$$D^* \xrightarrow{k_{rep}} p_{rep} + D_{T4}. \tag{6.38}$$

The rate k_{rep} must depend on the amount of gp45 present in the cell, such that

$$k_{rep} = \begin{cases} k_0 & \text{if } \; p_{45} = 0 \\ k_0 + k_{45}(p_{45}) & \text{if } \; p_{45} > 0 \end{cases} \tag{6.39}$$

where k_0 is the initial replication rate due only to p_{rep} and $k_{45}(p_{45})$ is the enhanced replication rate which is dependent on the amount of gp45 synthesised;

$$k_{45} = \frac{kp_{45}}{K_{45} + p_{45}}, \tag{6.40}$$

where k is the maximum rate of k_{45} and K_{45} is the half-maximal threshold of p_{45}.

Protein Synthesis	Transcription	Degradation	Ribosome Binding	Translation
E-M Transition Proteins	$b_{em} \xrightarrow{\omega_p} m_{em} + R_E$	$m_{em} \xrightarrow{d_m} \emptyset$	$r + m_{em} \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_{em}$	$n_{pg}a + c_{em} \xrightarrow{\nu_{em}} r + m_{em} + p_{em}$
RegB Proteins	$b_{reg} \xrightarrow{\omega_p} m_{reg} + R_E$	$m_{reg} \xrightarrow{d_m} \emptyset$	$r + m_{reg} \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_{reg}$	$n_{pg}a + c_{reg} \xrightarrow{\nu_{reg}} r + m_{reg} + p_{reg}$
Endolysins	$b_{end} \xrightarrow{\omega_p} m_{end} + R_E$	$m_{end} \xrightarrow{d_m} \emptyset$	$r + m_{end} \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_{end}$	$n_{pg}a + c_{end} \xrightarrow{\nu_{end}} r + m_{end} + p_{end}$
Gp45	$b_{45} \xrightarrow{\omega_p} m_{45} + R_M$	$m_{45} \xrightarrow{d_m} \emptyset$	$r + m_{45} \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_{45}$	$n_{pg}a + c_{45} \xrightarrow{\nu_{45}} r + m_{45} + p_{45}$
M-L Transition Proteins	$b_{ml} \xrightarrow{\omega_p} m_{ml} + R_M$	$m_{ml} \xrightarrow{d_m} \emptyset$	$r + m_{ml} \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_{ml}$	$n_{pg}a + c_{ml} \xrightarrow{\nu_{ml}} r + m_{ml} + p_{ml}$
Structural Proteins	$b_S \xrightarrow{\omega_p} m_S + R_L$	$m_S \xrightarrow{d_m} \emptyset$	$r + m_S \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_S$	$n_{pg}a + c_S \xrightarrow{\nu_S} r + m_S + p_S$
Holins	$b_h \xrightarrow{\omega_p} m_h + R_L$	$m_h \xrightarrow{d_m} \emptyset$	$r + m_h \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_h$	$n_{pg}a + c_h \xrightarrow{\nu_h} r + m_h + p_h$
Lysozyme	$b_{ly} \xrightarrow{\omega_p} m_{ly} + R_L$	$m_{ly} \xrightarrow{d_m} \emptyset$	$r + m_{ly} \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_{ly}$	$n_{pg}a + c_{ly} \xrightarrow{\nu_{ly}} r + m_{ly} + p_{ly}$
DNA Replication Proteins	$b_{rep} \xrightarrow{\omega_p} m_{rep} + R_{E,M,L}$	$m_{rep} \xrightarrow{d_m} \emptyset$	$r + m_{rep} \underset{k_u}{\overset{k_b}{\rightleftharpoons}} c_{rep}$	$n_{pg}a + c_{rep} \xrightarrow{\nu_{rep}} r + m_{rep} + p_{rep}$
Transcriptional Control	RNAP Reformation	RNAP Binding		
Early Period	R_E	$R_E + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_{em} + D_{T4}$		
		$R_E + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_{reg} + D_{T4}$		
		$P_E + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_{end} + D_{T4}$		
Middle Period	$R_E + p_{em} \xrightarrow{k_{em}} R_M$	$R_M + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_{ml} + D_{T4}$		
		$R_M + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_{45} + D_{T4}$		
Late Period	$R_M + p_{ml} \xrightarrow{k_{ml}} R_L$	$R_L + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_h + D_{T4}$		
		$R_L + DT4 \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_S + D_{T4}$		
		$R_L + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} b_{ly} + D_{T4}$		
		$P_{E,M,L} + D_{T4} \underset{k_{-t}}{\overset{k_t}{\rightleftharpoons}} T_{rep} + D_{T4}$		
DNA Replisome	DNA Binding	Replication		
	$p_{rep} + D_{T4} \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} D^* + D_{T4}$	$D^* \xrightarrow{k_{rep}} p_{rep} + D_{T4}$		

Table 6.2 *Table showing set of reactions used in the phage-adapted model. The expression of each transcriptive period-specific protein depends on which conformation of RNAP is present in the cell at any given time.*

6.2.4 Complete List of Reactions, Equations, and Parameters in the Phage Model

I have condensed the reactions into table 6.2, and list the resulting ODEs below with $x \in \{E,\ M,\ L\}$ to denote early, middle, and late respectively. The table of reactions in the phage model is similar to that of the foundation model, however new columns need to be added for RNAP reconfiguration, binding, and DNA replication. A table of parameters included in the phage model can be seen in table 6.3.

Derived from the reactions are a set of coupled ODEs. Beginning with the mRNAs $\overrightarrow{m}$

$$\frac{d\mathbf{m_x}}{dt} = \omega_p(a)\mathbf{b_x} - (\lambda + d_m) + \nu_{\mathbf{x}}(\mathbf{c_x}, a) - k_b r\mathbf{m_x} + k_u\mathbf{c_x}, \tag{6.41}$$

where $\mathbf{m_x} = [\overrightarrow{m_e}, \overrightarrow{m_m}, \overrightarrow{m_l}]$ and similarly for $\mathbf{b_x}$ and $\mathbf{c_x}$. That is, mRNAs are produced from transcription completing, translation completing, and translation complexes unbinding (terms one, three, and five) and removed by growth and

degradation, and translation complex formation (terms two and four).

The transcription complexes that are formed by RNAP molecules binding to DNA follow

$$\frac{d\mathbf{b_x}}{dt} = k_t R_x D_{T4} - k_{-t}\mathbf{b_x} - \omega_p \mathbf{b_x} - \lambda \mathbf{b_x}, \tag{6.42}$$

where k_t and k_{-t} are the rates of binding and unbinding of R_x to D_{T4}. Transcriptional complexes are generated from binding and removed by spontaneous unbinding, complete transcription, and dilution by growth.

RNAP reconfiguration by binding with the temporal switching proteins is expressed as

$$\frac{dR_e}{dt} = -k_{em} R_e p_{em} - k_t D_{T4} R_e + k_{-t}\overrightarrow{b_e} D_{T4} + \omega_p \overrightarrow{b_e} - \lambda R_e, \tag{6.43}$$

for RNAP with early-affinity;

$$\frac{dR_m}{dt} = k_{em} R_e p_{em} - k_{ml} R_m p_{ml} - k_t D_{T4} R_m + k_{-t}\overrightarrow{b_m} D_{T4} + \omega_p \overrightarrow{b_m} - \lambda R_m, \tag{6.44}$$

for RNAP with middle-affinity; and

$$\frac{dR_l}{dt} = k_{ml} R_m p_{ml} - k_t D_{T4} R_l + k_{-t}\overrightarrow{b_x} + \omega_p \overrightarrow{b_x} D_{T4} - \lambda R_l, \tag{6.45}$$

for RNAP with late-affinity.

The transition proteins em and ml are produced in accordance with

$$\frac{dp_{em}}{dt} = -k_{em} R_e + \nu_{em} c_{em} - \lambda p_{em} \tag{6.46}$$

and

$$\frac{dp_{ml}}{dt} = -k_{ml} R_m + \nu_{ml} c_{ml} - \lambda p_{ml}; \tag{6.47}$$

and the rest of the proteins are simply translated and diluted as

$$\begin{aligned}
\frac{dp_{reg}}{dt} &= \nu_{reg}c_{reg} - \lambda p_{reg}, \\
\frac{dp_{end}}{dt} &= \nu_{end}c_{end} - \lambda p_{end}, \\
\frac{dp_S}{dt} &= \nu_S c_S - \lambda p_S, \\
\frac{dp_l}{dt} &= \nu_l c_l - \lambda p_l, \\
\text{and} & \\
\frac{dp_h}{dt} &= \nu_h c_h - \lambda p_h
\end{aligned} \tag{6.48}$$

The replication of T4 DNA follows the ODE

$$\frac{dD_{T4}}{dt} = -k_d p_{rep} D_{T4} + k_{-d} D^* + k_{rep} D^* N_{T4}^2 - \lambda D_{T4}, \tag{6.49}$$

where the rates k_d and k_{-d} are the rates of binding and unbinding of p_{rep} to D_{T4} respectively and the rate k_{rep} is the rate of production of D_{T4} from a single replication complex D^*. The constant N_{T4} is the length of the T4 genome. The replication proteins p_{rep} thus follow

$$\frac{dp_{rep}}{dt} = -k_d p_{rep} D_{T4} + k_{-d} D^* + k_{rep} D^* N_{T4} + \nu_{rep} c_{rep} - \lambda p_{rep}, \tag{6.50}$$

and the replication complexes D^* follow

$$\frac{dD^*}{dt} = k_d p_{rep} D_{T4} - k_{rep} D^* N_{T4} - \lambda D^*. \tag{6.51}$$

The phage model differs from the foundational model equations in that DNA is modeled as an explicit species, and so transcription and replication have their own complexes. The rates of transcription must therefore be updated for the phage model. The expression for the rate of transcription ω_p is thus

$$\omega_p(\mathbf{b_x}, a) = \frac{\mathbf{b_x}}{N_p} \cdot \frac{\mathbf{w_x} a}{\theta_x + a}, \tag{6.52}$$

where N_p is the average length of a T4 gene; and the expression for the DNA replication rate k_{rep} is

$$k_{rep}(D^*, a) = \frac{D^*}{N_{T4}} \cdot \frac{k_{rmax} a}{K_r + a}, \tag{6.53}$$

Variable	Definition	Proposed Value	Units	
w_e	Max. Early Elongation Rate	100	bp s^{-1}	[85]
w_m	Max. Middle Elongation Rate	100	bp s^{-1}	[85]
w_l	Max. Late Elongation Rate	100	bp s^{-1}	[85]
k_{rmax}	Max. DNA Elongation Rate	1000	bp s^{-1}	[83]
k_{1t}	T_x Binding Rate	1	$molec^{-1}min^{-1}$	†
k_{-1t}	T_x Unbinding Rate		$molec^{-1}min^{-1}$	⋆
k_{em}	P_m Binding Rate	1	$molec^{-1}min^{-1}$	†
k_{ml}	P_l Binding Rate	1	$molec^{-1}min^{-1}$	†
k_{1d}	D^* Binding Rate	1	$molec^{-1}min^{-1}$	†
k_{-1d}	D^* Unbinding Rate		$molec^{-1}min^{-1}$	⋆
θ_e	Early Half-Maximal Threshold		molec	⋆
θ_m	Middle Half-Maximal Threshold		molec	⋆
θ_l	Late Half-Maximal Threshold		molec	⋆
K_r	Replication Half-Maximal Threshold		molec	⋆
N_{T4}	Length of T4 Genome	170,000	bp	[85]
N_p	Av. Length of T4 Genes	566	bp	*
Γ_{end}	Host Transcription Decay Rate		molec min^{-1}	⋆
Γ_{reg}	Early Transcription Decay Rate		molec min^{-1}	⋆
P_h	Initial Host RNAP Number	5000	molecs	

Table 6.3 *Table of definitions and proposed values for each rate variable. † based off the 2015 model. * a rough calculation based on the number of genes and the length of the genome. The three transcription elongation rates are not suggested to be numerically different, however the exact nature is not known and so three variables are kept. The half-maximal thresholds however will likely be different and only knowable through parameter sampling. ⋆ will be determined through parameter sampling.*

where k_{rmax} is the maximum rate of replication and K_r is a half-maximal threshold.

6.3 Discussion and Concluding Remarks

This model was produced during a time of total lab closure, and represents an effort to understand the functional dependencies that underpin the physiology of a T4-infected *E. coli*. It pays particular attention to the temporal control process of switching between transcriptive periods, and accounts for a singular host that is growing exponentially at time of infection. There exists no code for this model as the labs re-opened as the mathematics were finalised. It would not have been written without the pandemic.

As discussed at the beginning of the chapter, this model provides a basis for

predicting the effects of changes in the transcription and biosynthesis rates. In this way it could be used to show the expected effects of different antibiotics (potentially useful for combined therapy), environment conditions (nutrient density, temperature), and potentially even some bacterial immune systems that attack certain phage components. Extending the model to other phages and bacteria may be quite time-consuming, however with the baseline presented here may not be too complex to be done.

In conclusion, this model has the potential to be highly informative, but has limited immediate use as it has no code nor numerical results as a result of the constraints put on the work at the time. In order to fully realise the model, several experiments should be carried out. Namely, the transcriptive periods should be assessed via a similar experiment to Luke *et al* [78]. Additionally studies of protein expression and overall DNA content can be performed or data from existing studies collated. The model should be tightly fitted to experimental data, and those parameters that cannot be recovered from experiments can be parametrically fitted. There is also the distinct possibility that parts of the model can be trimmed down to form so-called 'toy-models', which focus particularly on one or two functional dependencies to inform experiments. These are much faster to design and write, and the mathematics presented here is detailed enough to form the basis of several different such models.

Chapter 7

Conclusions and Future Work

To sum up the main findings of this thesis I present a collection of conclusions in the order that their relevant chapters appear. To the extent of understanding the phenomenon of an apparent loss of volume of *E. coli* cells following infection by T7 phage a method for the measurement of single cell volume has been developed. The protocol makes use of a machine-learning analysis pipeline DeLTA2, developed by Dunlop *et al* to quickly and efficiently characterise the area of cells adsorbed to a glass cover slip, which is a proxy for cell volume. As DeLTA2 does not necessarily require fluorescent or otherwise marked cells or phage, this is a promising first look into a potential label-free method of observing phage infection at the single-cell level.

The developed protocol yielded measurements of the volume dips. On average infected cells lost 6.6%±0.02% of their volume when infected by T7 phage, and took on average 5.6±0.6 minutes to recover this lost volume. Over all cells and adjusting for noise there was approximately a 43%±5% chance to observe a dip in any given cell. It was found that these dips have a dependence on the temperature history of the bacteria, and those that had been loaded into the measurement chambers at room temperature were approximately 1.6 times more likely to display a dip than those loaded at 37°C. Additionally, these colder cells had approximately double the time-to-recover of their warmer counterparts.

Supporting the experimental data is a model that has been adapted from a hypoosmotic shock model that modeled the solute and volume loss in the case of a solute-permeable hole being opened in a cell membrane. The model predicts that the time-to-recover of the cell volume must be longer than the hole-open

duration of the hole. It also predicts that there should be a loss of volume in all cases of infection, if the assumption that the hole mediating solute loss is formed at time of infection is true. Therefore the implication is that if one increases the resolution of one's measurements, more dips should be observed. In the case of a non-solute-permeable hole opening, the volume of the cell is not predicted to change.

The mechanism by which temperature affects the cells in this way is not known. Presented in Chapter 5 is a discussion of various hypothesis, of which the most likely appears to be a physical effect wherein the membrane undergoes a partial phase-transition from liquid to gel at the colder temperature. This transition would have the effect of reducing the speed at which the lipid membrane can reorganise around the infection site, thereby increasing the duration of the hole and therefore the magnitude of the dip.

Additionally, a model of the T4 infection of a single *E. coli* cell has been mathematically formulated. This model pays specific attention to the temporal control of the infection, wherein T4 switches between three distinct transcriptive periods in a process mediated by T4-encoded gene products. This model has the potential to be highly informative of the effect of changes in the biosynthesis rates in the bacterium on the overall infection process and efficacy. Thus, the model could at its most complete provide valuable insights into the interaction between different antibiotics, environmental conditions, and cell conditions and the overall phage infection. In addition several toy models can be created from the mathematics presented, which could be used to further understanding of certain functional dependencies within the wider context.

7.1 Future Work

A clear avenue of future work is to use labelled phage or phage DNA to determine whether or not the losses in volume do truly correspond to the point of DNA entry to the cell. Additionally, these experiments can be repeated in different environmental conditions and with different antibiotics present. These would have the effect of changing certain known aspects of bacterial physiology, and therefore provide a more complete understanding of the origin of the volume dips, which here remains an open question. Additionally, if the variation in measurement of dips is due to membrane properties, being able to perform AFM

around an infection site would provide valuable insights into the stochasticity of the process.

Further to this, optimization of the baseline protocol developed in this thesis is required for future experiments to yield good information. Primarily, the segmentation error in the analysis should be minimized as much as possible with the tool used being switched away from DeLTA2 if necessary. Another large issue with the protocol is its low-throughput nature. Cell adhesion should be investigated for improvement to increase the number of analysable cells per FOV and different stabilisation methods should be used for the microscope. A perfect focus system would allow the stage to be stabilised in Z for good area measurements while being free in X and Y so that multiple FOVs can be taken from one slide. This would drastically reduce the required number of experiments per condition and enable many more conditions to be examined.

Once the dips in T7 are more understood, it would then be sensible to move to other phage. Each phage has different injection dynamics, and so while the dip may be more of a physical process than a biological one, this may change the visibility. If the technique can be refined for more than one phage such that a dip is observable in the majority of cases, it presents an exciting label-free time-precise marker of phage infection of a single cell. In order to support future developments, the full set of raw data can be provided on request, along with the various python scripts used to visualise it.

The model presented in Chapter 6 should in the first instance be evaluated for worth as a trimmed down toy model. Bringing the entire thing into computational functionality is a large undertaking, however as highlighted could be quite valuable. Experimental data should be collected so that the model can be parametrically fitted, and care taken to avoid local minima.

The videos used to produce the results presented in Chapter 5 can be found on Edinburgh Datashare in the Biophysics collection under the author's name. Any enquiries for clarification around said data can be directed to the author at hockingj4 at gmail.com, or to the principle supervisor Dr Aidan Brown at aidan.brown at ed.ac.uk.

Appendix A

Mathematical Description of Edge-Detection Methods Trialed in Chapter 3

Initially I picked three of the standard edge detection protocols that MatLab's Image Analysis Toolbox function 'edge' recognises: Sobel, Prewitt, and Canny. At the time I was not aware that Canny is essentially an extended methodology that uses the Sobel algorithm, merely that it was computationally more expensive. As such, I will describe these techniques in the above order despite settling for Canny detection. MatLab's default edge function is Sobel, and so that was my starting point. The Sobel, or Sobel-Feldman, Operator is an isotropic 3 x 3 image gradient operator initially presented in 1968 by authors Irwin Sobel and Gary Feldman[121]. It estimates the gradient of an image at a point by the vector sum of the four possible central gradient estimates in a 3 x 3 neighbourhood. For a 3 x 3 neighbourhood each simple gradient estimate is a sum of two orthogonal vectors. They are both directional derivative estimations multiplied by a unit vector specifying the derivative's direction. The sum of these four simple estimations multiplied by a unit vector specifying the derivative's direction. The sum of these four simple estimates is a vector sum of the 8 directional derivative vectors. For a point on a Cartesian grid e and its eight neighbours

$$p = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \tag{A.1}$$

the magnitude of the directional derivative estimate g for a given neighbour is

$$|g| = \left\langle \frac{value\ difference}{distance\ to\ neighbour} \right\rangle. \tag{A.2}$$

The direction of g will be the unit vector to the appropriate neighbour. The neighbours group into antipodal pairs: (a, i), (b, h), (c, g), and (f, g). Summing the estimates within each pair causes all the central 'e' values to cancel (as it is in the centre) leaving the following for the gradient estimate,

$$G = \frac{(c-g)}{4} * [1,1] + \frac{(a-i)}{4} * [-1,1] + \frac{(b-h)}{2} * [0,1] + \frac{(f-d)}{2} * [1,0], \tag{A.3}$$

which simplifies to

$$G = \left[\frac{(c-g-a+i)}{4} + \frac{(f-d)}{2}, \frac{(c-g+a-i)}{4} + \frac{(b-h)}{2} \right]. \tag{A.4}$$

In order to get the average gradient, one would normally divide the result by four. However, Sobel instead multiplies by four in order to 'preserve low order bits' as operations are often done in fixed point on small integers and divisions will lose these bits to rounding. The fourfold multiplication represents a double shift to the left, and leaves us with an estimate that is 16 times the average gradient. One can represent $G' = 4G$ as weighting functions for x and y components

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \tag{A.5}$$

which are referred to as the Sobel kernels. The edge finding operation convolves these kernels with the sample image in a sliding window convolution (figure, explain?) to produce a mask which is highly non-zero in areas where there is high amplitude change in pixel values (i.e. an edge) in the direction aligned with the gradient of each G_x and G_y (i.e. G_x will highlight edges in the horizontal

direction).

Prewitt follows a similar vein, instead using kernels

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \qquad G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}. \tag{A.6}$$

Convolution of an image with both the Sobel and Prewitt kernels is essentially the matrix equivalent of taking their first derivative. In principle Prewitt is the simplest example, where Sobel is the summation of the Prewitt kernel with a simple matrix where

$$G_x = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad G_y = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}. \tag{A.7}$$

In this way the diagonal edges are given a lower weight than the edges that lie along the cardinal directions.

Using only kernel convolution to find edges is generally considered to be quite unsophisticated due to the handling of noise. It does find all edges, weak or strong, in an image, and where there is zero or near-zero noise this is adequate. However I found that a simple convolution was not able to properly highlight cells within my ROI1s due to noise in the image(figure). In order to resolve this, I turned to the third option: Canny. The Canny algorithm functions by finding edges by gradient operator and then selecting only the strong edge in a thresholding process. In two dimensions the Canny method is a five-step process to define real, strong, edges in the image:

Canny Edge Detection

1. Smooth image by convolution with Gaussian matrix.
2. Apply Sobel-Feldman detection to smoothed image to find intensity gradients.
3. Remove spurious gradients by thresholding.
4. Apply a double threshold to locate weak and strong edges.

5. Track found edges with hysteresis. Suppress weak edges that are not connected to strong edges.

The first step comprises a sliding window convolution of the image with a Gaussian matrix e.g.

$$G = \frac{1}{273}\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}, \tag{A.8}$$

to blur the image. Blurring the image reduces the prominence of all edges in the image, highlighting the stronger edges, while reducing noise in the image. Any dimension of Gaussian kernel can be used, the larger it is the less sensitive the detector will be to noise, however the localization error in detecting edges will increase slightly with the dimension size of the kernel. For this reason a 5x5 matrix is generally the largest that one typically uses. The following step applies the operators in (2.5) to the image to generate the absolute gradient magnitudes in the x- and y-directions for each pixel. The edge strength is then given by the simple formula

$$|G| = |G_x| + |G_y|. \tag{A.9}$$

In addition to the edge strength the edge direction is required. It is given trivially by

$$\theta = tan^{-1}\left(\frac{G_y}{G_x}\right), \tag{A.10}$$

and rounded to one of four angles: 0° for horizontal, 90° for vertical, and 45° and 135° for the two diagonals. The reasoning behind this quantization is clear if one considers a pixel **a**:

$$\begin{matrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & a & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{matrix} \tag{A.11}$$

one can see that there are only those four directions that an edge can exist along for that pixel. As an example, any edge direction equal to 22° would be set to 0° for the description of edges. Once the direction and magnitude are known, the algorithm moves to the fourth step with the process of strong edge selection.

First, a gradient magnitude thresholding method is applied to thin the edges out, then a pair of thresholds are applied to the thinned mask to filter out edges that are weak and unconnected. For each pixel in the gradient mask the edge strength of that pixel is compared to that of the pixels in the positive and negative gradient directions. If the current pixel has a higher edge strength than other pixels in the mask with the same direction the value is preserved, otherwise it is suppressed. I.e. a pixel with a vertical gradient across it will be compared in strength to the pixels above and below it, if it is the strongest of those in the comparison then its value is preserved, if not it is suppressed. After the edge mask has been thinned one is left with a relatively accurate representation of edges within an image. However some spurious edges remain due to noise and colour variation. Thus a low threshold and high threshold are applied to the edge mask. If a pixel gradient magnitude is higher than the high threshold it is a *strong* edge pixel, and if the value is in between the two thresholds it is marked as a *weak* edge pixel. Pixel values below the low threshold are suppressed. The fifth and final step to the process is to track edges by hysteresis. This is necessary as, while one knows that the strong edges ought to be included in the edge mask, one needs to be able to decide which weak edges to include. An 8-connectivity blob analysis is applied, wherein subsets of connected components in a neighbourhood of 8 pixels are labelled as belonging to each other, and weak edges that are not connected to one or more strong edge pixels are suppressed.

Bibliography

[1] Abedon, S. T. "Lysis from without." *Bacteriophage* 1.

[2] Addy, H. S., A. Askora, T. Kawasaki, M. Fujie, and T. Yamada. "Loss of virulence of the phytopathogen Ralstonia solanacearum through infection by ϕ RSM filamentous phages." *Phytopathology* 102, 5: (2012) 469–477.

[3] Alemohammad, M. M., and C. J. Knowles. "Osmotically Induced Volume and Turbidity Changes of Escherichia colidue to Salts, Sucrose and Glycerol, with Particular Reference to the Rapid Permeation of Glycerol into the Cell." *Microbiology* 82, 1: (1974) 125–142. https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-82-1-125.

[4] Anderson, T. "The reactions of bacterial viruses with their host cells." .

[5] Annunziato, A. "DNA Packaging: Nucleosomes and Chromatin." *Nature Education* 1, 1.

[6] Bartual, S. G., J. M. Otero, C. Garcia-Doval, A. L. Llamas-Saiz, R. Kahn, G. C. Fox, and M. J. van Raaij. "Structure of the bacteriophage T4 long tail fiber receptor-binding tip." *Proceedings of the National Academy of Sciences* 107, 47: (2010) 20,287–20,292. https://www.pnas.org/doi/abs/10.1073/pnas.1011218107.

[7] Baslé, A., G. Rummel, P. Storici, J. P. Rosenbusch, and T. Schirmer. "Crystal Structure of Osmoporin OmpC from E. coli at 2.0 Å." *Journal of Molecular Biology* 362, 5: (2006) 933–942. https://www.sciencedirect.com/science/article/pii/S0022283606009855.

[8] Bayer, M. E. "Adsorption of bacteriophages to adhesions between wall and membrane of Escherichia coli." *J. Virol.* 2, 4: (1968) 346–356.

[9] Bertozzi Silva, J., Z. Storms, and D. Sauvageau. "Host receptors for bacteriophage adsorption." *FEMS Microbiology Letters* 363, 4: (2016) 1–11.

[10] Bialecka-Fornal, M., H. J. Lee, H. A. DeBerg, C. S. Gandhi, and R. Phillips. "Single-cell census of mechanosensitive channels in living bacteria." *PLoS One* 7, 3: (2012) e33,077.

[11] Boeckman, J., A. Korn, G. Yao, A. Ravindran, C. Gonzalez, and J. Gill. "Sheep in wolves' clothing: Temperate T7-like bacteriophages and the origins of the Autographiviridae." *Virology* 568: (2022) 86–100. https://www.sciencedirect.com/science/article/pii/S0042682222000204.

[12] Bossi, L., J. A. Fuentes, G. Mora, and N. Figueroa-Bossi. "Prophage Contribution to Bacterial Population Dynamics." *Journal of Bacteriology* 185, 21: (2003) 6467–6471. https://journals.asm.org/doi/abs/10.1128/JB.185.21.6467-6471.2003.

[13] Broeker, N. K., and S. Barbirz. "Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection." *Molecular Microbiology* 105, 3: (2017) 353–357. https://onlinelibrary.wiley.com/doi/abs/10.1111/mmi.13729.

[14] Buda, R., Y. Liu, J. Yang, S. Hegde, K. Stevenson, F. Bai, and T. Pilizota. "Dynamics of ¡i¿Escherichia coli¡/i¿’s passive response to a sudden decrease in external osmolarity." *Proceedings of the National Academy of Sciences* 113, 40: (2016) E5838–E5846. https://www.pnas.org/doi/abs/10.1073/pnas.1522185113.

[15] Burkhardt, R. W. "Lamarck, evolution, and the inheritance of acquired characters." *Genetics* 194, 4: (2013) 793–805.

[16] Cayley, D. S., H. J. Guttman, and M. T. Record. "Biophysical characterization of changes in amounts and activity of Escherichia coli cell and compartment water and turgor pressure in response to osmotic stress." *Biophysical Journal* 78, 4: (2000) 1748–1764.

[17] Chang, C.-Y., P. Kemp, and I. J. Molineux. "Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell." *Virology* 398, 2: (2010) 176–186.

[18] Cieślik, M., N. Bagińska, E. Jończyk-Matysiak, A. Wegrzyn, G. Wegrzyn, and A. Górski. "Temperate Bacteriophages—The Powerful Indirect Modulators of Eukaryotic Cells and Immune Functions." *Viruses* 13, 6. https://www.mdpi.com/1999-4915/13/6/1013.

[19] Conley, M. P., and W. B. Wood. "Bacteriophage T4 whiskers: a rudimentary environment sensing device." *Proceedings of the National Academy of Sciences of the United States of America* 72, 9: (1975) 3701–3705.

[20] Cooper, S., and C. E. Helmstetter. "Chromosome replication and the division cycle of Escherichia coli Br." *Journal of molecular biology* 31, 3: (1968) 519–540.

[21] Cottrell, M. T., and C. A. Suttle. "Dynamics of lytic virus infecting the photosynthetic marine picoflagellate Micromonas pusilla." *Limnology and Oceanography* 40, 4: (1995) 730–739.

[22] Cuervo, A., M. Fàbrega-Ferrer, C. Machón, J. J. Conesa, F. J. Fernández, R. Pérez-Luque, M. Pérez-Ruiz, J. Pous, M. C. Vega, J. L. Carrascosa, and M. Coll. "Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism." *Nature Communications* 10, 1: (2019) 3746. https://doi.org/10.1038/s41467-019-11705-9.

[23] De Paepe, M., and F. Taddei. "Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages." *PLoS Biol.* 4, 7: (2006) e193.

[24] Delbruck, M. "the Growth of Bacteriophage and Lysis of the Host." *The Journal of General Physiology* 23, 5: (1940) 643–660.

[25] Deng, Y., M. Sun, and J. W. Shaevitz. "Direct Measurement of Cell Wall Stress Stiffening and Turgor Pressure in Live Bacterial Cells." *Phys. Rev. Lett.* 107: (2011) 158,101. https://link.aps.org/doi/10.1103/PhysRevLett.107.158101.

[26] Depping, R., C. Lohaus, H. E. Meyer, and W. Rüger. "The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: Target proteins identified." *Biochemical and Biophysical Research Communications* 335, 4: (2005) 1217–1223.

[27] Dick, D. A. T. *Cell Water*. 1966.

[28] Donachie, W. D. "Relationship between Cell Size and Time of Initiation of DNA replication." *Nature* 219: (1968) 1077–1079.

[29] Dy, R. L., R. Przybilski, K. Semeijn, G. P. C. Salmond, and P. C. Fineran. "A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism." *Nucleic Acids Research* 42, 7: (2014) 4590–4605.

[30] Edgar, R., A. Rokney, M. Feeney, S. Semsey, M. Kessel, M. B. Goldberg, S. Adhya, and A. B. Oppenheim. "Bacteriophage infection is targeted to cellular poles." *Molecular Microbiology* 68, 5: (2008) 1107–1116. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2008.06205.x.

[31] Edwards, A. W. "The Genetical Theory of Natural Selection." *Genetics* 154, 4: (2000) 1419–1426.

[32] Evilevitch, A., L. Lavelle, C. M. Knobler, E. Raspaud, and W. M. Gelbart. "Osmotic pressure inhibition of DNA ejection from phage." *Proceedings of the National Academy of Sciences of the United States of America* 100, 16: (2003) 9292 – 9295. Cited by: 256; All Open Access, Green Open Access.

[33] Farré, A., F. Marsà, and M. Montes-Usategui. "Optimized back-focal-plane interferometry directly measures forces of optically trapped particles." *Opt. Express* 20, 11: (2012) 12,270–12,291. https://opg.optica.org/oe/abstract.cfm?URI=oe-20-11-12270.

[34] Frixione, E., and L. Ruiz-Zamarripa. "The "scientific catastrophe" in nucleic acids research that boosted molecular biology." *Journal of Biological Chemistry* 294, 7: (2019) 2249–2255.

[35] Fruciano, E. "Phage as an antimicrobial agent: d'Herelle's heretical theories and their role in the decline of phage prophylaxis in the West." 18, 1: (2007) 19–26.

[36] Garro, A. J., and J. Marmur. "Defective bacteriophages." *Journal of Cellular Physiology* 76, 3: (1970) 253–263. https://onlinelibrary.wiley.com/doi/abs/10.1002/jcp.1040760305.

[37] Ge, P., D. Scholl, P. G. Leiman, X. Yu, J. F. Miller, and Z. H. Zhou. "Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states." *Nature Structural and Molecular Biology* 22, 5: (2015) 377–382.

[38] González-Huici, V., M. Salas, and J. M. Hermoso. "The push–pull mechanism of bacteriophage Ø29 DNA injection." *Molecular Microbiology* 52, 2: (2004) 529–540. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2004.03993.x.

[39] Görke, B., and J. Stülke. "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients." *Nature Reviews Microbiology* 6, 8: (2008) 613–624. https://doi.org/10.1038/nrmicro1932.

[40] Grinius, L., and R. Daugelavičius. "Depolarization of Escherichia coli cytoplasmic membrane by bacteriophages T4 and lambda: evidence for induction of ion-permeable channels." *Journal of Electroanalytical Chemistry* 253, 2: (1988) 235–245.

[41] Hadas, H., M. Einav, and A. Zaritsky. "Development Depends on the Physiology." *Journal Of Bacteriology* , 1 997: (1994) 179–185.

[42] Hazel, J. R. "Thermal Adaptation in Biological Membranes: Is Homeoviscous Adaptation the Explanation?" *Annual Review of Physiology* 57, 1: (1995) 19–42. https://doi.org/10.1146/annurev.ph.57.030195.000315. PMID: 7778864.

[43] Hegazy, Y. A., C. M. Fernando, and E. J. Tran. "The balancing act of R-loop biology: The good, the bad, and the ugly." *J. Biol. Chem.* 295, 4: (2020) 905–913.

[44] Henning, U., B. Höhn, and I. Sonntag. "Cell Envelope and Shape of Escherichia coli K12." *European Journal of Biochemistry* 39, 1: (1973) 27–36. https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1432-1033.1973.tb03099.x.

[45] Herendeen, S. L., R. A. VanBogelen, and F. C. Neidhardt. "Levels of major proteins of Escherichia coli during growth at different temperatures." *J. Bacteriol.* 139, 1: (1979) 185–194.

[46] Hershey, A. D., and M. Chase. "INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE." *The Journal of General Physiology* 36, 1: (1952) 39–56. http://jgp.rupress.org/content/36/1/39.

[47] Hinton, D. M. "Transcriptional control in the prereplicative phase of T4 development." *Virology Journal* 7, 1: (2010) 289. http://www.virologyj.com/content/7/1/289.

[48] Hoffmann, A., B. Bukau, and G. Kramer. "Structure and function of the molecular chaperone Trigger Factor." *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1803, 6: (2010) 650–661. https://www.sciencedirect.com/science/article/pii/S0167488910000303. Molecular Chaperones and Intracellular Protein Transport.

[49] Holmes, C. M., M. Ghafari, A. Abbas, V. Saravanan, and I. Nemenman. "Luria-Delbrück, revisited: The classic experiment does not rule out Lamarckian evolution." *Physical Biology* 14, 5.

[50] Hu, B., W. Margolin, I. J. Molineux, and J. Liu. "The bacteriophage t7 virion undergoes extensive structural remodeling during infection." *Science* 339, 6119: (2013) 576–579.

[51] ———. "Structural remodeling of bacteriophage T4 and host membranes during infection initiation." *Proceedings of the National Academy of Sciences* 112, 35: (2015) E4919–E4928.

[52] Huang, J., J. Villemain, R. Padilla, and R. Sousa. "Mechanisms by which T7 lysozyme specifically regulates T7 RNA polymerase during different phases of transcription 11Edited by R. Ebright." *Journal of Molecular Biology* 293, 3: (1999) 457–475. https://www.sciencedirect.com/science/article/pii/S0022283699931352.

[53] Hyman, P., and S. Abedon. "Bacteriophage (overview)." In *Encyclopedia of Microbiology (Third Edition)*, edited by Moselio Schaechter, Oxford: Academic Press, 2009, 322–338. Third edition edition. https://www.sciencedirect.com/science/article/pii/B9780123739445000201.

[54] Islam, M. Z., A. Fokine, M. Mahalingam, Z. Zhang, C. Garcia-Doval, M. J. van Raaij, M. G. Rossmann, and V. B. Rao. "Molecular anatomy of the receptor binding module of a bacteriophage long tail fiber." *PLOS Pathogens* 15, 12: (2019) 1–21. https://doi.org/10.1371/journal.ppat.1008193.

[55] Jeembaeva, M., M. Castelnovo, F. Larsson, and A. Evilevitch. "Osmotic Pressure: Resisting or Promoting DNA Ejection from Phage?" *Journal of Molecular Biology* 381, 2: (2008) 310–323. https://www.sciencedirect.com/science/article/pii/S0022283608006682.

[56] Jones, P. G., and M. Inouye. "The cold-shock response–a hot topic." *Mol. Microbiol.* 11, 5: (1994) 811–818.

[57] Josslin, R. "The Lysis Mechanism Affecting of Phage Lysis." *Virology* 40: (1970) 719–726.

[58] Kanamaru, S., Y. Ishiwata, T. Suzuki, M. G. Rossmann, and F. Arisaka. "Control of Bacteriophage T4 Tail Lysozyme Activity During the Infection Process." *Journal of Molecular Biology* 346, 4: (2005) 1013–1020. https://www.sciencedirect.com/science/article/pii/S0022283604016122.

[59] Kashlev, M., E. Nudler, A. Goldfarb, T. White, and E. Kutter. "Bacteriophage T4 Alc protein: A transcription termination factor sensing local modification of DNA." *Cell* 75, 1: (1993) 147–154.

[60] Kemp, P., M. Gupta, and I. J. Molineux. "Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism." *Molecular Microbiology* 53, 4: (2004) 1251–1265.

[61] Kiro, R., S. Molshanski-Mor, I. Yosef, S. L. Milam, H. P. Erickson, and U. Qimron. "Gene product 0.4 increases bacteriophage T7 competitiveness by inhibiting host cell division." *Proceedings of the National Academy of Sciences* 110, 48: (2013) 19,549–19,554. https://www.pnas.org/doi/abs/10.1073/pnas.1314096110.

[62] Koch, A. L. "Shrinkage of growing Escherichia coli cells by osmotic challenge." *Journal of Bacteriology* 159, 3: (1984) 919–924. https://journals.asm.org/doi/abs/10.1128/jb.159.3.919-924.1984.

[63] Köhler, A. "New Method of Illumination for Photomicrographical Purposes." *Journal of the Royal Microscopical Society* 261–262.

[64] Kosaka, Y., W. Aoki, M. Mori, S. Aburaya, Y. Ohtani, H. Minakuchi, and M. Ueda. "Selected reaction monitoring for the quantification of Escherichia coli ribosomal proteins." *PLOS ONE* 15, 12: (2020) 1–14. https://doi.org/10.1371/journal.pone.0236850.

[65] Krasnopeeva, E., C.-J. Lo, and T. Pilizota. "Single-Cell Bacterial Electrophysiology Reveals Mechanisms of Stress-Induced Damage." *Biophysical Journal* 116, 12: (2019) 2390–2399. https://www.sciencedirect.com/science/article/pii/S0006349519303923.

[66] Kreuzer, K. N., and J. R. Brister. "Initiation of bacteriophage T4 DNA replication and replication fork dynamics: A review in the Virology Journal series on bacteriophage T4 and its relatives." *Virology Journal* 7, 1: (2010) 358. http://www.virologyj.com/content/7/1/358.

[67] Kropinski, A. M., A. Mazzocco, T. E. Waddell, E. Lingohr, and R. P. Johnson. *Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay*, Totowa, NJ: Humana Press, 2009, 69–76. https://doi.org/10.1007/978-1-60327-164-6_7.

[68] Krüger, D. H., and C. Schroeder. "Bacteriophage T3 and bacteriophage T7 virus-host cell interactions." *Microbiological Reviews* 45, 1: (1981) 9–51. https://journals.asm.org/doi/abs/10.1128/mr.45.1.9-51.1981.

[69] Kuhn, A. H., H. Jütte, and E. Kellenberger. "Involvement of the bacterial groM gene product in bacteriophage T7 reproduction. II. A reduced level of ion concentrations causes the blockage of T7 maturation in K-12-M cells." *Journal of Virology* 47, 3: (1983) 540–552. https://journals.asm.org/doi/abs/10.1128/jvi.47.3.540-552.1983.

[70] Kumar, P., and A. Libchaber. "Pressure and temperature dependence of growth and morphology of Escherichia coli: experiments and stochastic model." *Biophys. J.* 105, 3: (2013) 783–793.

[71] Labedan, B., and E. B. Goldberg. "Requirement for membrane potential in injection of phage T4 DNA." *Proceedings of the National Academy of Sciences* 76, 9: (2006) 4669–4673.

[72] Leforestier, A., S. Brasilès, M. de Frutos, E. Raspaud, L. Letellier, P. Tavares, and F. Livolant. "Bacteriophage T5 DNA Ejection under Pressure." *Journal of Molecular Biology* 384, 3: (2008) 730–739. http://dx.doi.org/10.1016/j.jmb.2008.09.035.

[73] Lerman, L. S. "A transition to a compact form of DNA in polymer solutions." *Proc. Natl. Acad. Sci. U. S. A.* 68, 8: (1971) 1886–1890.

[74] Lerouge, I., and J. Vanderleyden. "O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions." *FEMS Microbiology Reviews* 26, 1: (2002) 17–47. https://doi.org/10.1111/j.1574-6976.2002.tb00597.x.

[75] Li, Z., J. Wu, and Z.-G. Wang. "Osmotic Pressure and Packaging Structure of Caged DNA." *Biophysical Journal* 94: (2008) 737–746.

[76] Liu, D. "Escherichia coli." In *Encyclopedia of Microbiology (Fourth Edition)*, edited by Thomas M. Schmidt, Oxford: Academic Press, 2019, 171–182. Fourth edition edition. https://www.sciencedirect.com/science/article/pii/B9780128012383022911.

[77] Lugagne, J.-B., H. Lin, and M. J. Dunlop. "DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning." *PLOS Computational Biology* 16, 4: (2020) 1–18. https://doi.org/10.1371/journal.pcbi.1007673.

[78] Luke, K., A. Radek, X. P. Liu, J. Campbell, M. Uzan, R. Haselkorn, and Y. Kogan. "Microarray analysis of gene expression during bacteriophage T4 infection." *Virology* 299, 2: (2002) 182–191.

[79] Luria, S. E., and M. Delbrück. "MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE." *Genetics* 28, 6: (1943) 491–511. https://www.genetics.org/content/28/6/491.

[80] Löf, D., K. Schillén, B. Jönsson, and A. Evilevitch. "Forces Controlling the Rate of DNA Ejection from Phage ." *Journal of Molecular Biology* 368, 1: (2007) 55–65. https://www.sciencedirect.com/science/article/pii/S0022283607001349.

[81] Maghsoodi, A., A. Chatterjee, I. Andricioaei, and N. C. Perkins. "Dynamic Model Exposes the Energetics and Dynamics of the Injection Machinery for Bacteriophage T4." *Biophysical Journal* 113, 1: (2017) 195–205. https://www.sciencedirect.com/science/article/pii/S0006349517305684.

[82] Martínez, E., and J. Campos-Gómez. " Pf Filamentous Phage Requires UvrD for Replication in Pseudomonas aeruginosa ." *mSphere* 1, 1: (2016) 1–14.

[83] McCarthy, D., C. Minner, H. Bernstein, and C. Bernstein. "DNA elongation rates and growing point distributions of wild-type phage T4 and a DNA-delay amber mutant." *Journal of Molecular Biology* 106, 4: (1976) 963–981.

[84] Mika, J. T., G. Van Den Bogaart, L. Veenhoff, V. Krasnikov, and B. Poolman. "Molecular sieving properties of the cytoplasm of Escherichia coli and consequences of osmotic stress." *Molecular Microbiology* 77, 1: (2010) 200–207. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2010.07201.x.

[85] Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Ruger. "Bacteriophage T4 Genome." *Microbiology and Molecular Biology Reviews* 67, 1: (2003) 86–156.

[86] Moak, M., and I. J. Molineux. "Peptidoglycan hydrolytic activities associated with bacteriophage virions." *Mol. Microbiol.* 51, 4: (2004) 1169–1183.

[87] Moffat, B., and F. Studier. *Cell* 49: (1987) 221 – 227. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85080841808&partnerID=40&md5=35fc3bde46b8c2ebfc12bb07ef51009e. Cited by: 0.

[88] Moffatt, B. A., and F. W. Studier. "Entry of bacteriophage T7 DNA into the cell and escape from host restriction." *J. Bacteriol.* 170, 5: (1988) 2095–2105.

[89] Molineux, I. J. "The T7 Group." In *The Bacteriophages Second Edition*, edited by R. Calendar, Oxford: Oxford University Press, 2006, chapter 20, 277–301.

[90] ———. "No syringes please, ejection of phage T7 DNA from the virion is enzyme driven." *Molecular Microbiology* 40, 1: (2001) 1–8. https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2001.02357.x.

[91] ———. "Fifty-three years since Hershey and Chase; Much ado about pressure but which pressure is it?" *Virology* 344, 1: (2006) 221–229.

[92] Molineux, I. J., and D. Panja. "Popping the cork: mechanisms of phage genome ejection." 11, March: (2013) 194–204.

[93] Molshanski-Mor, S., I. Yosef, R. Kiro, R. Edgar, M. Manor, M. Gershovits, M. Laserson, T. Pupko, and U. Qimron. "Revealing bacterial targets of growth inhibitors encoded by bacteriophage T7." *Proc. Natl. Acad. Sci. U. S. A.* 111, 52: (2014) 18,715–18,720.

[94] Moussatova, A., C. Kandt, M. L. O'Mara, and D. P. Tieleman. "ATP-binding cassette transporters in Escherichia coli." *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1778, 9: (2008) 1757–1771. `https://www.sciencedirect.com/science/article/pii/S0005273608001946`. Structural proteomics of the cell envelope of Gram-negative bacteria.

[95] Mulder, E., and C. L. Woldringh. "Plasmolysis bays in Escherichia coli: are they related to development and positioning of division sites?" *Journal of Bacteriology* 175, 8: (1993) 2241–2247. `https://journals.asm.org/doi/abs/10.1128/jb.175.8.2241-2247.1993`.

[96] Nabergoj, D., P. Modic, and A. Podgornik. "Effect of bacterial growth rate on bacteriophage population growth rate." *MicrobiologyOpen* 7, 2: (2018) 1–10.

[97] Neidhardt, F. C., P. L. Bloch, and D. F. Smith. "Culture medium for enterobacteria." *J. Bacteriol.* 119, 3: (1974) 736–747.

[98] Nobrega, F. L., M. Vlot, P. A. de Jonge, L. L. Dreesens, H. J. E. Beaumont, R. Lavigne, B. E. Dutilh, and S. J. J. Brouns. "Targeting mechanisms of tailed bacteriophages." *Nature Reviews Microbiology* 16.

[99] Orsini, G., S. Igonet, C. Pène, B. Sclavi, M. Buckle, M. Uzan, and A. Kolb. "Phage T4 early promoters are resistant to inhibition by the anti-sigma factor AsiA." *Molecular Microbiology* 52, 4: (2004) 1013–1028. `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2004.04038.x`.

[100] O'Connor, O. M., R. N. Alnahhas, J.-B. Lugagne, and M. J. Dunlop. "DeLTA 2.0: A deep learning pipeline for quantifying single-cell spatial and temporal dynamics." *PLOS Computational Biology* 18, 1: (2022) 1–18. `https://doi.org/10.1371/journal.pcbi.1009797`.

[101] Panja, D., and I. J. Molineux. "Dynamics of bacteriophage genome ejection in vitro and in vivo." *Phys. Biol.* 7, 4: (2010) 045,006.

[102] Parfitt, T. "Georgia: an unlikely stronghold for bacteriophage therapy." *The Lancet* 365, 9478: (2005) 2166–2167. `https://doi.org/10.1016/S0140-6736(05)66759-1`.

[103] Pfeifer, E., R. A. Bonnin, and E. P. C. Rocha. "Phage-Plasmids Spread Antibiotic Resistance Genes through Infection and Lysogenic Conversion." *mBio* 13, 5: (2022) e01,851–22. https://journals.asm.org/doi/abs/10.1128/mbio.01851-22.

[104] Pilizota, T., and J. W. Shaevitz. "Origins of Escherichia coli growth rate and cell shape changes at high external osmolality." *Biophys. J.* 107, 8: (2014) 1962–1969.

[105] Pilizota, T., and J. Shaevitz. "Plasmolysis and Cell Shape Depend on Solute Outer-Membrane Permeability during Hyperosmotic Shock in E. coli." *Biophysical Journal* 104, 12: (2013) 2733–2742. https://www.sciencedirect.com/science/article/pii/S0006349513005614.

[106] Publications service. "On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux." *Research in Microbiology* 158, 7: (2007) 553–554. https://www.sciencedirect.com/science/article/pii/S0923250807001507. Commemoration of Felix Herelle's discovery.

[107] Qiu, X., D. C. Rau, V. A. Parsegian, L. T. Fang, C. M. Knobler, and W. M. Gelbart. "Salt-dependent DNA-DNA spacings in intact bacteriophage λ reflect relative importance of DNA self-repulsion and bending energies." *Phys. Rev. Lett.* 106, 2: (2011) 028,102.

[108] Rakhuba, D. V., E. I. Kolomiets, E. S. Dey, and G. I. Novik. "Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell." *Pol. J. Microbiol.* 59, 3: (2010) 145–155.

[109] Rao, V. B., and L. W. Black. "Structure and assembly of bacteriophage T4 head." *Virology Journal* 7, 1: (2010) 356. https://doi.org/10.1186/1743-422X-7-356.

[110] Record, M. T., Jr, E. S. Courtenay, D. S. Cayley, and H. J. Guttman. "Responses of E. coli to osmotic stress: large changes in amounts of cytoplasmic solutes and water." *Trends Biochem. Sci.* 23, 4: (1998) 143–148.

[111] Ronneberger, O., P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation.", 2015.

[112] Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse. "A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase." *Science* 262, 5138: (1993) 1407–1413.

[113] Salmond, G. P. C., and P. C. Fineran. "A century of the phage: past, present and future." *Nature Reviews Microbiology* 13, 12: (2015) 777–786. https://doi.org/10.1038/nrmicro3564.

[114] Savalia, D., W. Robins, S. Nechaev, I. Molineux, and K. Severinov. "The role of the T7 Gp2 inhibitor of host RNA polymerase in phage development." *J. Mol. Biol.* 402, 1: (2010) 118–126.

[115] Secor, P. R., J. M. Sweere, L. A. Michaels, A. V. Malkovskiy, D. Lazzareschi, E. Katznelson, J. Rajadas, M. E. Birnbaum, A. Arrigoni, K. R. Braun, S. P. Evanko, D. A. Stevens, W. Kaminsky, P. K. Singh, W. C. Parks, and P. L. Bollyky. "Filamentous bacteriophage promote biofilm assembly and function." *Cell Host and Microbe* 18, 5: (2015) 549–559. http://dx.doi.org/10.1016/j.chom.2015.10.013.

[116] Serwer, P., E. T. Wright, K. W. Hakala, and S. T. Weintraub. "Evidence for bacteriophage T7 tail extension during DNA injection." *BMC Research Notes* 1, 1: (2008) 36. https://doi.org/10.1186/1756-0500-1-36.

[117] Shabala, L., J. Bowman, J. Brown, T. Ross, T. McMeekin, and S. Shabala. "Ion transport and osmotic adjustment in Escherichia coli in response to ionic and non-ionic osmotica." *Environmental Microbiology* 11, 1: (2009) 137–148. https://ami-journals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2008.01748.x.

[118] Shimada, T., Y. Yamazaki, K. Tanaka, and A. Ishihama. "The Whole Set of Constitutive Promoters Recognized by RNA Polymerase RpoD Holoenzyme of Escherichia coli." *PLOS ONE* 9, 3: (2014) 1–23. https://doi.org/10.1371/journal.pone.0090447.

[119] Singh, M. I., B. Ganesh, and V. Jain. "On the domains of T4 phage sliding clamp gp45: An intermolecular crosstalk governs structural stability and biological activity." *Biochimica et Biophysica Acta (BBA) - General Subjects* 1861, 1, Part A: (2017) 3300–3310. https://www.sciencedirect.com/science/article/pii/S0304416516302938.

[120] Smith, D. E., D. L. Anderson, C. Bustamante, S. J. Tans, S. Grimes, and S. B. Smith. "The bacteriophage ϕ29 portal motor can package DNA against a large internal force." *Nature* 413, 6857: (2002) 748–752. www.nature.com.

[121] Sobel, I., and G. Feldman. "A 3 x 3 Isotropic Gradient Operator for Image Processing.", 1968. A talk at the Stanford Artificial Project.

[122] Stone, R. "Stalin ' s Forgotten Cure." *Science* 298: (2002) 728–731.

[123] Storms, Z. J., and D. Sauvageau. "Modeling tailed bacteriophage adsorption: Insight into mechanisms." *Virology* 485: (2015) 355–362. http://dx.doi.org/10.1016/j.virol.2015.08.007.

[124] Studier, F. "The genetics and physiology of bacteriophage T7." *Virology* 39, 3: (1969) 562–574. https://www.sciencedirect.com/science/article/pii/0042682269901044.

[125] Summers, W. C. "Bacteriophage therapy." *Annu. Rev. Microbiol.* 55, 1: (2001) 437–451.

[126] Svircev, A., D. Roach, and A. Castle. "Framing the future with bacteriophages in agriculture." *Viruses* 10, 5: (2018) 1–13.

[127] Tabib-Salazar, A., B. Liu, A. Shadrin, L. Burchell, Z. Wang, Z. Wang, M. G. Goren, I. Yosef, U. Qimron, K. Severinov, S. J. Matthews, and S. Wigneshweraraj. "Full shut-off of Escherichia coli RNA-polymerase by T7 phage requires a small phage-encoded DNA-binding protein." *Nucleic Acids Research* 45, 13: (2017) 7697–7707. https://doi.org/10.1093/nar/gkx370.

[128] Taslem Mourosi, J., A. Awe, W. Guo, H. Batra, H. Ganesh, X. Wu, and J. Zhu. "Understanding Bacteriophage Tail Fiber Interaction with Host Surface Receptor: The Key ldquo;Blueprintrdquo; for Reprogramming Phage Host Range." *International Journal of Molecular Sciences* 23, 20. https://www.mdpi.com/1422-0067/23/20/12146.

[129] Terradot, G., E. Krasnopeeva, P. S. Swain, and T. Pilizota. "The proton motive force determines Escherichia coli's robustness to extracellular pH." *bioRxiv* https://www.biorxiv.org/content/early/2021/11/19/2021.11.19.469321.

[130] Thomans, P., G. Terradot, V. Danos, and A. Weiße. "Sources, propagation and consequences of stochasticity in cellular growth." *Nature Communications* 9, 4528.

[131] Tsujimoto, H., N. Gotoh, and T. Nishino. "Diffusion of macrolide antibiotics through the outer membrane of Moraxella catarrhalis." *Journal of Infection and Chemotherapy* 5, 4: (1999) 196–200. https://www.sciencedirect.com/science/article/pii/S1341321X99713450.

[132] Twort, F. "AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES." *The Lancet* 186, 4814: (1915) 1241–1243. https://www.sciencedirect.com/science/article/pii/S0140673601203833. Originally published as Volume 2, Issue 4814.

[133] Walker, P. J., S. G. Siddell, E. J. Lefkowitz, A. R. Mushegian, E. M. Adriaenssens, P. Alfenas-Zerbini, D. M. Dempsey, B. E. Dutilh, M. L. García, R. Curtis Hendrickson, S. Junglen, M. Krupovic, J. H. Kuhn, A. J. Lambert, M. Łobocka, H. M. Oksanen, R. J. Orton, D. L. Robertson, L. Rubino, S. Sabanadzovic, P. Simmonds, D. B. Smith, N. Suzuki, K. Van Doorslaer, A.-M. Vandamme, A. Varsani, and F. M. Zerbini. "Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022)." *Archives of Virology* 167, 11: (2022) 2429–2440. https://doi.org/10.1007/s00705-022-05516-5.

[134] Walkinshaw, M., P. Taylor, S. Sturrock, C. Atanasiu, T. Berge, R. Henderson, J. Edwardson, and D. Dryden. "Structure of Ocr from Bacteriophage T7, a Protein that Mimics B-Form DNA." *Molecular Cell* 9, 1: (2002) 187–194. https://www.sciencedirect.com/science/article/pii/S1097276502004355.

[135] Wang, X., L. Li, Y. Shao, J. Wei, R. Song, S. Zheng, Y. Li, and F. Song. "Effects of the Laplace pressure on the cells during cytokinesis." *iScience* 24, 9: (2021) 102,945. https://www.sciencedirect.com/science/article/pii/S2589004221009135.

[136] Wang, Y.-K., E. Krasnopeeva, S.-Y. Lin, F. Bai, T. Pilizota, and C.-J. Lo. "Comparison of Escherichia coli surface attachment methods for single-cell microscopy." *Scientific Reports* 9, 1: (2019) 19,418. https://doi.org/10.1038/s41598-019-55798-0.

[137] Washizaki, A., T. Yonesaki, and Y. Otsuka. "Characterization of the interactions between Escherichia coli receptors, LPS and OmpC, and bacteriophage T4 long tail fibers." *MicrobiologyOpen* 5, 6: (2016) 1003–1015.

[138] Weiße, A. Y., D. A. Oyarzún, V. Danos, and P. S. Swain. "Mechanistic links between cellular trade-offs, gene expression, and growth." *Proceedings of the National Academy of Sciences* 112, 9: (2015) E1038–E1047. https://www.pnas.org/doi/abs/10.1073/pnas.1416533112.

[139] Wietzorrek, A., H. Schwarz, C. Herrmann, and V. Braun. "The genome of the novel phage Rtp, with a rosette-like tail tip, is homologous to the genome of phage T1." *J. Bacteriol.* 188, 4: (2006) 1419–1436.

[140] Wilkens, K., B. Tiemann, F. Bazan, and W. Rüger. "ADP-Ribosylation and early transcription regulation by bacteriophage T4." In *ADP-Ribosylation in Animal Tissues*, edited by Friedrich Haag, and Friedrich Koch-Nolte, Plenum Press, 1977, chapter 8, 71–82.

[141] William Studier, F. "Relationships among different strains of T7 and among T7-related bacteriophages." *Virology* 95, 1: (1979) 70–84. https://www.sciencedirect.com/science/article/pii/0042682279904021.

[142] Wood, J. M. "Osmosensing by Bacteria." *Science's STKE* 2006, 357: (2006) pe43–pe43. https://www.science.org/doi/abs/10.1126/stke.3572006pe43.

[143] Xu, H., X. Bao, W. Hong, A. Wang, K. Wang, H. Dong, J. Hou, R. Govinden, B. Deng, and H. Y. Chenia. "Biological Characterization and Evolution of Bacteriophage T7-holin During the Serial Passage Process." *Frontiers in Microbiology* 12. https://www.frontiersin.org/articles/10.3389/fmicb.2021.705310.

[144] Yang, H., Y. Ma, Y. Wang, H. Yang, W. Shen, and X. Chen. "Transcription regulation mechanisms of bacteriophages." *Bioengineered* 5, 5: (2014) 300–304. https://doi.org/10.4161/bioe.32110. PMID: 25482231.

[145] Yap, M. L., and M. G. Rossmann. "Structure and function of bacteriophage T4." *Future Microbiology* 9, 12: (2014) 1319–1337.

[146] Ye, F., I. Kotta-Loizou, M. Jovanovic, X. Liu, D. T. Dryden, M. Buck, and X. Zhang. "Structural basis of transcription inhibition by the DNA mimic protein Ocr of bacteriophage T7." *eLife* 9: (2020) e52,125. https://doi.org/10.7554/eLife.52125.

[147] Yoo, J.-H., and U. L. RajBhandary. "Requirements for translation reinitiation in Escherichia coli: roles of initiator tRNA and initiation factors IF2 and IF3." *Mol. Microbiol.* 67, 5: (2008) 1012–1026.

[148] Zárybnický, V. "Mechanism of T-even DNA ejection." *Journal of Theoretical Biology* 22, 1: (1968) 33–42.

[149] Zimmerman, S., and L. Murphy. "Macromolecular crowding and the mandatory condensation of DNA in bacteria." *FEBS Letters* 390: (1996) 245–248.